

การพยากรณ์ระยะทางการปั่นจักรยานเพื่อสุขภาพโดยใช้วิธีซัพพอร์ตเวกเตอร์
แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม

FORECASTING A DISTANCE OF CYCLING FOR HEALTH USING
A HYBRID GA-SVR APPROACH



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาคณิตศาสตร์ประยุกต์
ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2564

KMITL-2021-SC-M-001-059

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

FORECASTING A DISTANCE OF CYCLING FOR HEALTH USING
A HYBRID GA-SVR APPROACH



A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE
DEGREE OF MASTER OF SCIENCE IN APPLIED MATHEMATICS
DEPARTMENT OF MATHEMATICS SCHOOL OF SCIENCE
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2021

KMITL-2021-SC-M-001-059

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2021

SCHOOL OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การพยากรณ์ระยะทางการปั่นจักรยานเพื่อสุขภาพโดยใช้วิธีซัพพอร์ต เวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม
ชื่อนักศึกษา	นางสาวพรวิภา มีเคลือบ
รหัสนักศึกษา	62605013
ปริญญา	วิทยาศาสตรมหาบัณฑิต (คณิตศาสตร์ประยุกต์)
ภาควิชา	คณิตศาสตร์
ปีการศึกษา	2564
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.บุษยมาส พิมพ์พรรณชาติ

บทคัดย่อ

การศึกษานี้เกี่ยวข้องกับ การพยากรณ์ระยะทางการปั่นจักรยานที่เหมาะสมในเวลา 30 นาที สำหรับบุคคลที่มีข้อจำกัดทางด้านร่างกาย โดยเฉพาะอย่างยิ่งผู้ที่มีโรคประจำตัวหรือผู้สูงอายุซึ่งต้องให้ความสำคัญเป็นพิเศษ มีการรวบรวมข้อมูลโดยการทำแบบสอบถามผู้ที่ปั่นจักรยานที่สนามปั่นจักรยานในเขตลาดกระบัง จังหวัดกรุงเทพมหานคร จำนวน 94 คน สำหรับงานวิจัยนี้สนใจการพยากรณ์ระยะทางที่เหมาะสมด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมเพื่อปรับปรุงความแม่นยำและได้เสนอการเปรียบเทียบประสิทธิภาพของแบบจำลองด้วยวิธีการวิเคราะห์การถดถอยเชิงพหุคูณและวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยพิจารณาจากค่าความผิดพลาดด้วยรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยและเปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย ผลลัพธ์ที่ได้คือวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมเป็นวิธีที่แม่นยำที่สุดในการพยากรณ์ระยะทางการปั่นจักรยาน

คำสำคัญ : การปั่นจักรยาน; การออกกำลังกายเพื่อสุขภาพ; วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย; ขั้นตอนวิธีเชิงพันธุกรรม; บุคคลที่มีข้อจำกัด

Title	Forecasting a Distance of Cycling for Health using a Hybrid GA-SVR Approach
Student	Miss Pawnwipa Meeklueb
Student ID	62605013
Degree	Master of Science (Applied Mathematics)
Department	Mathematics
Academic Year	2021
Advisor	Asst. Prof. Dr. Busayamas Pimpunchat

ABSTRACT

The aim of this study is to forecast the optimal cycling distance in 30 minutes approximately for physically constrained individuals, especially those having a congenital disease or elderly person. Data illustrations are collected from having 94 people who cycled in Ladkrabang by questionnaires. The research focuses on forecasting optimal distances by exploiting Genetic Algorithm-Based Support Vector Regression (GA-SVR) that will be employed to improve the accuracy of the numerical study. We offer performance comparisons of our model against Multiple Linear Regression (MLR) and Support Vector Regression (SVR) algorithms. The experimental results demonstrate that GA-SVR outperforms more than MLR and SVR based on the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). This GA-SVR model is proven to be an effective approach to forecast cycling distances.

Keyword: Cycling; Health Exercise; Support Vector Regression; Genetic Algorithm; Physically Constrained Individuals.

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก ผศ.ดร.บุษยมาส พิมพ์พรรณชาติ อาจารย์ที่ปรึกษา ที่ได้ให้คำแนะนำการใช้โปรแกรม แนวคิดทางการศึกษาค้นคว้าหาความรู้ และตลอดจนแก้ไขข้อบกพร่องต่าง ๆ มาโดยตลอด วิทยานิพนธ์เล่มนี้เสร็จสมบูรณ์ ผู้ศึกษาจึงขอกราบขอบพระคุณเป็นอย่างสูงและขอบพระคุณคณาจารย์ทุกท่านที่ได้กรุณาประสิทธิ์ประสาทวิชาความรู้อันถือเป็นพื้นฐานที่สำคัญยิ่งในการทำวิจัยครั้งนี้ ขอขอบคุณทางคณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังที่มอบทุนการศึกษาให้ผู้ศึกษา ขอบพระคุณ คุณพ่อ คุณแม่ ที่ให้กำลังใจในการศึกษา และให้การสนับสนุนในการทำวิจัยตลอดมา คุณค่า และประโยชน์อันพึงมีส่งเป็นผลจากงานวิจัยนี้ ขอมอบแต่ คุณพ่อ คุณแม่ ญาติพี่น้อง ครู อาจารย์ และผู้มีพระคุณทุกท่าน

พรวิภา มีเคลือบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญรูป	VII
บทที่ 1 บทนำ	1
1.1 ความเป็นมาและความสำคัญ	1
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ขั้นตอนการศึกษา	3
1.5 ประโยชน์ที่คาดว่าจะได้รับ	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 การพัฒนาสุขภาพ	4
2.1.1 ด้านร่างกาย	4
2.1.2 ด้านจิตใจ	5
2.1.3 ด้านสติปัญญา	5
2.1.4 ด้านสังคม	5
2.2 โครงข่ายประสาทเทียม	7
2.2.1 วิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM)	8
2.2.2 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SVR)	14
2.3 ขั้นตอนวิธีเชิงพันธุกรรม (GA)	17
2.3.1 กระบวนการขั้นตอนวิธีเชิงพันธุกรรม	18
2.4 การวัดความถูกต้องของการพยากรณ์	20
2.4.1 รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE)	20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และเผยแพร่ไปยังเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
2.4.2 เปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)	20
2.5 งานวิจัยที่เกี่ยวข้อง	20
บทที่ 3 วิธีดำเนินการวิจัย	24
3.1 ข้อมูลที่นำมาใช้	24
3.1.1 การสร้างเครื่องมือที่ใช้ในการวิจัย	25
3.2 ขั้นตอนการสร้างแบบจำลอง	28
3.2.1 การวิเคราะห์การถดถอยเชิงพหุคูณ	28
3.2.1.1 ทำการนำเข้าข้อมูลจากไฟล์ Excel	29
3.2.2 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย	30
3.2.3 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม	30
3.3 แผนงานวิจัย	32
บทที่ 4 ผลการวิจัยและอภิปรายผล	33
4.1 ผลการวิจัยและการอภิปรายการพยากรณ์ระยะทางที่เหมาะสมในการปั่นจักรยาน	33
4.1.1 ผลลัพธ์ที่ได้จากโปรแกรมวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม	34
4.2 ผลการวิจัยและการอภิปรายผลหลังจากมีการปรับปรุงค่าพารามิเตอร์	36
4.2.1 สรุปผล	38
บทที่ 5 สรุปผลการวิจัย	39
5.1 สรุปผลวิจัย	39
เอกสารอ้างอิง	41
ภาคผนวก ก จักรยานอยู่กับที่อ้างอิงระดับปริญญาตรี	44
ภาคผนวก ข ข้อมูลที่นำมาใช้ในงานวิจัย	46
ภาคผนวก ค ผลงานวิจัยที่ได้รับการตีพิมพ์	50
ประวัติผู้เขียน	63

สารบัญตาราง

ตารางที่	หน้า
ตารางที่ 2.1 ระบบไหลเวียนโลหิตและระบบหายใจ	4
ตารางที่ 2.2 ศัพท์เฉพาะในวิธีเชิงพันธุกรรม	18
ตารางที่ 3.1 ตัวอย่างข้อมูล	27
ตารางที่ 4.1 แสดงผลเปรียบเทียบ ทั้ง 3 วิธี	35
ตารางที่ 4.2 แสดงค่าพารามิเตอร์ที่เหมาะสม	38
ตารางที่ 4.3 แสดงผลการวัดค่าคลาดเคลื่อนกับเวลาที่ใช้ในการคำนวณ	38



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่	หน้า
รูปที่ 2.1 ตัวอย่างหน้าจอที่แสดงข้อมูลในเครื่องบินจักรยานอยู่กับที่	7
รูปที่ 2.2 แสดงตัวอย่างของเส้นระนาบเกินในการแบ่งข้อมูลออกเป็น 2 คลาส	8
รูปที่ 2.3 แสดงเส้นระยะขอบและเวกเตอร์ซัพพอร์ต	9
รูปที่ 2.4 แสดงสมการระนาบเกิน	9
รูปที่ 2.5 แสดงการแบ่งคลาสที่ $y_{i+} = 1$ และ $y_{i-} = -1$	11
รูปที่ 2.6 แสดงปัญหาขอบแข็ง(Hard margin)	12
รูปที่ 2.7 แสดงปัญหาขอบอ่อน(Soft margin)	12
รูปที่ 2.8 แสดงตัวอย่างการเปรียบเทียบระหว่างการกำหนดค่า C มากและค่า C น้อย	13
รูปที่ 2.9 แสดงรูปแบบการแบ่งกลุ่มด้วย Kernel ต่างๆโดย 1) Linear Kernel 2) Polynomial Kernel 3) Radial Basis Function (RBF) Kernel 4) Sigmoid Kernel	14
รูปที่ 2.10 นกฟินช์ (Finch) ตามภูมิภาคต่างๆ (อ้างอิง: www.sites.google.com)	17
รูปที่ 2.11 แสดงขั้นตอนวิธีเชิงพันธุกรรม	19
รูปที่ 3.1 แสดง Package ที่นำเข้า	28
รูปที่ 3.2 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรต่างๆ	29
รูปที่ 3.3 แสดงการนำเข้าข้อมูลในโปรแกรม R	29
รูปที่ 3.4 แสดงข้อมูลระยะทางจริง	30
รูปที่ 3.5 แสดงผลลัพธ์จากวิธี SVR	30
รูปที่ 3.6 แผนงานวิจัย	32
รูปที่ 4.1 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี MLR	33
รูปที่ 4.2 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี SVR	33
รูปที่ 4.3 แสดงการทำซ้ำ 10 รอบ	34
รูปที่ 4.4 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี GA-SVR	34
รูปที่ 4.5 เปรียบเทียบทั้ง 3 วิธี	35
รูปที่ 4.6 แสดงการทำซ้ำ 50 รอบ	36
รูปที่ 4.7 แสดงการทำซ้ำ 100 รอบ	36

สารบัญรูป(ต่อ)

รูปที่	หน้า
รูปที่ 4.8 แสดงการทำซ้ำ 500 รอบ	37
รูปที่ 4.9 เปรียบเทียบการทำซ้ำ 10 50 100 และ 500 รอบ	37



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และ VIII อ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

การศึกษาวิจัยเรื่อง การพยากรณ์ระยะทางของการปั่นจักรยานเพื่อสุขภาพโดยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR) ผู้วิจัยมีแนวคิดและแรงบันดาลใจเกี่ยวกับการสร้างต้นแบบการพยากรณ์ให้กับผู้ที่มีข้อจำกัดด้านสุขภาพให้รู้ถึงขีดจำกัดของตนเองเพื่อไม่ให้ออกกำลังกายมากเกินไปจนเกินไปจนอาจทำให้ร่างกายบาดเจ็บในภายหลัง จึงได้เขียนวัตถุประสงค์ของงานวิจัย กำหนดขอบเขตการวิจัย ตลอดจนคาดการณ์ถึงประโยชน์ที่จะได้รับจากการศึกษาวิจัยในครั้งนี้ และนำเสนอในหัวข้อดังกล่าว

1.1 ความเป็นมาและความสำคัญ

ทุกวันนี้ผู้คนเริ่มสนใจการออกกำลังกายมากขึ้น การออกกำลังกายมีหลายรูปแบบทั้งแบบมีอุปกรณ์และไม่มีอุปกรณ์ นอกจากนี้ยังมีทั้งข้อดีและข้อเสีย ประโยชน์ของการออกกำลังกายคือช่วยให้ร่างกายแข็งแรงแต่การออกกำลังกายมากเกินไปอาจทำให้เกิดการบาดเจ็บได้ ดังนั้นจึงควรรู้ขีดจำกัดการออกกำลังกายของตัวเองเพื่อลดความเสี่ยงที่จะเกิดผลเสียต่อร่างกายในภายหลัง การปั่นจักรยานเป็นการออกกำลังกายที่ดีสำหรับทุกเพศทุกวัย แต่ในบางกลุ่มที่มีข้อจำกัดในการออกกำลังกาย เช่น ผู้ที่เป็นโรคความดันโลหิตสูง ผู้สูงอายุที่เป็นโรคเบาหวาน หรือกลุ่มเสี่ยงที่จะเป็นโรคหลอดเลือดหัวใจ การออกกำลังกายที่เพียงพอควบคู่ไปกับการรับประทานอาหารที่มีประโยชน์สามารถลดความเสี่ยงในการเกิดโรคได้ [1, 2] โดยทั่วไป แนะนำให้ออกกำลังกายสำหรับผู้ใหญ่อายุ 18-64 ปี โดยออกกำลังกายแบบแอโรบิกระดับความเข้มข้นปานกลางอย่างน้อย 150 นาทีต่อสัปดาห์ หรือ 30 นาที ห้าครั้งต่อสัปดาห์ [3] ผู้ที่มีข้อจำกัดทางด้านร่างกายสามารถยืดหยุ่นได้โดยออกกำลังกาย 2-3 ครั้งต่อสัปดาห์ [4] ดังนั้นในงานวิจัยนี้ เราจึงสนใจระยะทางที่เหมาะสมที่สุดสำหรับการปั่นจักรยานในระยะเวลา 30 นาที เพื่อเป็นแนวทางในการหลีกเลี่ยงผลการออกกำลังกายที่มากเกินไปจนส่งผลเสียมากกว่าผลดี

การศึกษาก่อนหน้านี้ได้เสนอวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression : MLR) [5] ในการพยากรณ์โดยเก็บข้อมูลจากผู้ที่ปั่นจักรยานในสนามลู่วิ่งจักรยานมาเป็นกลุ่มตัวอย่างในการพยากรณ์ระยะทางการปั่นจักรยานอยู่กับที่ (Stationary bike) ในขณะที่นั้นยังไม่มีมีการแพร่ระบาดของเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ 2019 (2019-nCoV) ซึ่งในปัจจุบันเนื่องจากสถานการณ์การแพร่ระบาดของเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ 2019 มีความรุนแรงมากขึ้นทำให้การดำเนินชีวิตของทุกคนเปลี่ยนไปทำให้ผู้คนไม่สามารถไปออกกำลังกายในที่สาธารณะได้ การออกกำลังกายที่บ้านถือเป็นทางเลือกที่ปลอดภัย จากเว็บไซต์ Research And Markets [6] ซึ่งเป็นแหล่งข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชั้นนำของโลกสำหรับรายงานการวิจัยตลาดต่างประเทศและข้อมูลการตลาด กล่าวว่ายอดขายอุปกรณ์ฟิตเนสเติบโตขึ้น 170% ในช่วงล็อกดาวน์จากไวรัสโคโรนาสายพันธุ์ใหม่เนื่องจากยิมได้ปิดตัวลงทำให้ผู้คนสนใจซื้ออุปกรณ์ออกกำลังกายมากขึ้นเช่น exercise bikes, treadmills and rowing machines

สำหรับเทคนิคในการพยากรณ์แบบดั้งเดิมที่นิยมใช้มากที่สุดคือวิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression : MLR) เป็นแบบจำลองความสัมพันธ์ระหว่างตัวแปรอิสระกับสิ่งที่สนใจหรือตัวแปรตามและจะมีประสิทธิภาพสูงมากขึ้นหากสามารถหาตัวแปรอิสระที่สามารถอธิบายตัวแปรตามได้ครอบคลุม แต่ในการพยากรณ์แบบสมัยใหม่ได้สนใจศึกษาเกี่ยวกับวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression : SVR) เป็นวิธีที่มีแนวคิดโครงสร้างการทำงานคล้ายคลึงกับการทำงานของเซลล์สมอง มีความสามารถในการเรียนรู้ได้จากประสบการณ์ สามารถใช้ในการทำนายหรือพยากรณ์พฤติกรรมที่ไม่เชิงเส้นได้แต่การพยากรณ์จะมีประสิทธิภาพสูงก็ต่อเมื่อสามารถหาค่าพารามิเตอร์ทั้ง 3 ค่า (*Cost (C)* *Gamma (γ)* และ *Epsilon (ϵ)*) ได้อย่างเหมาะสม จากงานวิจัยมีหลายวิธีในการนำมาประยุกต์ใช้ค้นหาพารามิเตอร์ทั้ง 3 แต่ที่สนใจและนำมาศึกษาในงานวิจัยนี้คือวิธีเชิงพันธุกรรม (Genetic Algorithm : GA) เป็นวิธีที่เป็นไปตามหลักทางพันธุศาสตร์และการคัดเลือกทางธรรมชาติเป็นวิธีที่เหมาะสมในการนำมาใช้เพิ่มประสิทธิภาพของแบบจำลอง การพยากรณ์ด้วยวิธีที่กล่าวมาข้างต้นเหมาะสมที่จะนำมาใช้ในการศึกษาเปรียบเทียบค่า พยากรณ์ระยะทางการปั่นจักรยานที่เหมาะสมโดยวัดจากราคที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) และ เปอร์เซนต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

1.2 วัตถุประสงค์ของงานวิจัย

- 1) เพื่อวิเคราะห์ข้อมูลระยะทางการปั่นจักรยานเพื่อสุขภาพสำหรับผู้ที่มิชอบจำกัดทางด้านร่างกาย
- 2) เพื่อศึกษาและนำเสนอวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR) ในการพยากรณ์ระยะทางการปั่นจักรยาน
- 3) สามารถพยากรณ์ระยะทางการปั่นจักรยานที่เหมาะสมสำหรับในผู้ที่มิชอบจำกัดทางด้านร่างกายได้

1.3 ขอบเขตของงานวิจัย

- 1) กลุ่มตัวอย่างคือคนกรุงเทพฯ (เขตลาดกระบัง) ที่มีการออกกำลังกายโดยใช้จักรยาน ที่ไม่มีปัญหาเกี่ยวกับข้อเท้า และปัญหาเกี่ยวกับข้อเข่า ไม่มีโรคเฉพาะที่อาจมีผลต่อประสิทธิภาพการทำงาน

ของร่างกาย ไม่มีความบกพร่องทางการมองเห็น เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- 2) การใช้แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR) ในการพยากรณ์
- 3) ข้อมูลที่ใช้คือ เพศ ส่วนสูง น้ำหนัก ระยะทาง ความถี่ในการปั่นจักรยานครั้งต่อสัปดาห์ และอัตราการเต้นของหัวใจ
- 4) งานวิจัยนี้ใช้โปรแกรมภาษา R ในส่วนของการประมวลผลและการทดสอบประสิทธิภาพของแบบจำลอง

1.4 ขั้นตอนการศึกษา

- 1) ศึกษาวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SVR) และวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)
- 2) สร้างแบบจำลองโดยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SVR) และวิธีแมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)
- 3) เปรียบเทียบแบบจำลองจากค่าคาดเคลื่อนจากการหาค่ากลางที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และเปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (MAPE)
- 4) ปรับปรุงแบบจำลองด้วยวิธีแมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR) พยากรณ์ระยะทางการปั่นจักรยานที่เหมาะสม
- 5) ประเมินผลและสรุปผลการวิจัย

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทราบระยะทางการปั่นจักรยานที่เหมาะสมโดยประมาณสำหรับผู้ที่มีข้อจำกัดทางด้านร่างกายเพื่อไม่ให้เกิดอาการบาดเจ็บในภายหลังอันเนื่องมาจากการออกกำลังกายที่มากเกินไป
- 2) เป็นแนวทางสำหรับประกอบการตัดสินใจด้านการออกกำลังกายเพื่อสุขภาพสำหรับบุคคลที่มีข้อจำกัดของการออกกำลังกายด้วยการปั่นจักรยาน
- 3) เป็นประโยชน์ทางด้านอุตสาหกรรมสุขภาพ สามารถพัฒนาต่อได้โดยนำแบบจำลองนี้ไปสร้างเป็นโปรแกรมคำนวณระยะทางที่เหมาะสมในจักรยานอยู่กับที่
- 4) สำหรับบุคคลที่สนใจผู้วิจัยหวังว่าจะเป็นประโยชน์โดยสามารถนำไปเป็นแนวทางในการพยากรณ์ระยะทางการปั่นจักรยานเพื่อให้ความแม่นยำที่ดียิ่งขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้กล่าวถึงประโยชน์ของการออกกำลังกายและงานวิจัยทฤษฎีที่เกี่ยวข้องกับการสร้างแบบจำลองโดยใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถอดรอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)

2.1 การพัฒนาสุขภาพ

การออกกำลังกาย หมายถึง การประกอบกิจกรรมใด ๆ ที่ทำให้ร่างกายหรือส่วนต่าง ๆ ของร่างกายเกิดการเคลื่อนไหว และมีผลให้ระบบต่าง ๆ ของร่างกายเกิดความสมบูรณ์ แข็งแรงและทำงานได้อย่างมีประสิทธิภาพ ในการออกกำลังกายต้องออกกำลังกายอย่างถูกต้อง ตามหลักการทางวิทยาศาสตร์ และการฝึกกีฬาจึงจะเกิดประโยชน์ ซึ่งประโยชน์ของการออกกำลังกายสามารถแบ่งได้ดังนี้ [5]

2.1.1 ด้านร่างกาย

การออกกำลังกายจะทำให้หัวใจและระบบต่างๆ ของร่างกายมีการเปลี่ยนแปลงและพัฒนาไปในทางที่ดีขึ้น ดังนี้

- 1 ระบบกล้ามเนื้อ กล้ามเนื้อจะมีการเจริญเติบโตขึ้น ทำให้ขนาดของเส้นใยกล้ามเนื้อขยายใหญ่ขึ้น (Hypertrophy) ให้ประสิทธิภาพในการทำงานของกล้ามเนื้อเพิ่มขึ้น กล้ามเนื้อจะทำงานได้นานขึ้น เนื่องจากการใช้ออกซิเจนน้อย
- 2 ระบบโครงร่าง ในขณะที่ออกกำลังกายกระดูกจะถูกดึง ถูกบีบจากแรงกล้ามเนื้อ ซึ่งจะกระตุ้นให้กระดูกมีการเจริญเติบโตขึ้น
- 3 ระบบไหลเวียนโลหิตและระบบหายใจ การออกกำลังกายเป็นประจำทำให้ระบบไหลเวียนโลหิตและระบบหายใจทำงานได้อย่างมีประสิทธิภาพดังตารางที่ 2.1

ตารางที่ 2.1 ระบบไหลเวียนโลหิตและระบบหายใจ

ระบบไหลเวียนโลหิต	ระบบหายใจ
1. กล้ามเนื้อหัวใจมีขนาดเพิ่มขึ้น	1. ความสามารถในการจับออกซิเจนของร่างกาย(Maximum Oxygen uptake) มีค่าเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 ระบบไหลเวียนโลหิตและระบบหายใจ(ต่อ)

2. อัตราการเต้นของชีพจรในขณะที่พักลดลง รวมทั้งอัตราการเต้นของชีพจรขณะออกกำลังกายต่ำกว่าระดับสูงสุดข้างลง	2. การหายใจมีประสิทธิภาพมากขึ้น
3. การเพิ่มปริมาณเม็ดเลือดแดง และฮีโมโกลบิน	3. ความจุชีพ(Vital capacity) เพิ่มขึ้น
4. ช่วยลดไขมันในเส้นเลือด	4. ทรวงอกขยายใหญ่ขึ้น กล้ามเนื้อที่เกี่ยวกับการหายใจแข็งแรงขึ้น ความยืดหยุ่นของปอดเพิ่มขึ้น
5. เพิ่มประสิทธิภาพในการทำงานของหัวใจขณะพัก คือ อัตราการเต้นของหัวใจต่ำลง	5. ลดอันตรายของโรคที่เกี่ยวข้องกับระบบหายใจและระบบไหลเวียนโลหิต

2.1.2 ด้านจิตใจ

การออกกำลังกายอย่างสม่ำเสมอ จะส่งผลทำให้เป็นคนที่มีบุคลิกที่มั่นคงสามารถปรับตัวเข้ากับสิ่งแวดล้อมได้ดี มีมนุษยสัมพันธ์ที่ดีสามารถปรับตัวเมื่อได้รับความเครียดได้ดี มีความฉลาดทางอารมณ์ (EQ) ที่ในระดับดี

2.1.3 ด้านสติปัญญา

นอกจากการออกกำลังกายมีผลต่อด้านจิตใจแล้ว ผู้ที่ออกกำลังกายอย่างสม่ำเสมอ จะมีความสามารถในการแก้ปัญหา มีความสามารถในการคิดอย่างสร้างสรรค์ มีความสามารถในการเรียนรู้ มีความสามารถในการปรับตัวเข้ากับเหตุการณ์ต่างๆได้ดี

2.1.4 ด้านสังคม

การออกกำลังกายจะช่วยเพิ่มความมีวุฒิภาวะทางสังคม มีความฉลาดทางสังคม เนื่องการออกกำลังกายเป็นกิจกรรมที่ส่งเสริมให้คนมีปฏิสัมพันธ์มากขึ้น จนนำไปสู่การพัฒนาทักษะทางสังคมที่ดี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การออกกำลังกายด้วยการปั่นจักรยานปัจจุบันความนิยมเรื่องการบินจักรยานได้เพิ่มขึ้นสูงมากในสังคมไทย อาจเป็นเพราะมาจากปัญหาเรื่องการจราจรที่ติดขัด ราคาพลังงาน ที่สูงขึ้นเรื่อย ๆ มลภาวะทางอากาศที่เป็นพิษ และความต้องการในการออกกำลังกายเพื่อสุขภาพแต่ไม่มีเวลา เป็นต้น

การหันมาปั่นจักรยานถือว่าเป็นอีกทางเลือกหนึ่งที่สามารถตอบโจทย์ในการช่วยลดและแก้ปัญหาดังกล่าวได้เป็นอย่างดี โดยเฉพาะคนในกลุ่มวัยรุ่นและกลุ่มวัยทำงานที่ได้ให้ความสนใจกับการปั่นจักรยานเป็นอย่างมาก จนเกิดเป็นเทรนด์ขึ้นในสังคม ซึ่งเราจะสามารถสังเกตได้ว่าปัจจุบัน มีชมรมคนรักจักรยานเกิดขึ้น หรือว่าจะเป็นงานการกุศลในปัจจุบันที่ได้ให้ความสนใจในการเลือกใช้จักรยานมาเป็นกิจกรรมหลักในงานเพื่อดึงดูดให้คนมาร่วมงาน

การปั่นจักรยานจัดอยู่ในการออกกำลังกายแบบแอโรบิก (Aerobic Exercise) การออกกำลังกายประเภทนี้ถือว่าสำคัญต่อการทำงานของร่างกาย โดยช่วยให้อัตราการเต้นของหัวใจและการหายใจดีขึ้น ช่วยขยายผนังหลอดเลือด ลดความดันโลหิต เผาผลาญไขมันส่วนเกินในร่างกาย ลดระดับน้ำตาลในเลือด ลดอาการอักเสบ และเพิ่มระดับไขมันดี ทั้งนี้ การออกกำลังกายแบบแอโรบิกควบคู่กับการลดน้ำหนักยังช่วยลดระดับคอเลสเตอรอล ควรทำกิจกรรมต่อเนื่องกันอย่างน้อยวันละ 30 นาที หรือสัปดาห์ละ 150 นาที เช่น เดินเร็ว ว่ายน้ำ จ็อกกิ้ง ปั่นจักรยาน หรือทำกิจกรรมเข้าจังหวะ อย่างไรก็ตาม การออกกำลังกายแบบแอโรบิกควรอยู่ในขอบเขตที่เหมาะสม ไม่ควรหักโหมจนหายใจไม่ทัน เวียนศีรษะ เจ็บหรือแน่นหน้าอก หรือรู้สึกแสบร้อนกลางทรวงอก ทั้งนี้ ควรอบอุ่นร่างกายโดยไม่ควรเกิน 20 นาที [7] หรือคลายกล้ามเนื้อทุกครั้ง จิบน้ำระหว่างออกกำลังกายให้เพียงพอ ส่วนผู้ที่มีปัญหาสุขภาพบางอย่าง เช่น ป่วยเป็นโรคหัวใจวาย หรือโรคไต ควรจำกัดปริมาณของเหลวตามแพทย์สั่ง ไม่ควรดื่มน้ำขณะออกกำลังกายมากเกินไป อีกทั้งควรแต่งตัวให้เหมาะสมกับสภาพอากาศในกรณีที่ออกกำลังกายกลางแจ้ง และไม่ออกกำลังกายในที่ที่อากาศหนาวหรือร้อนเกินไป เนื่องจากอาจเป็นลมแดด หากอากาศร้อนมาก หรืออุณหภูมิร่างกายลดต่ำลงในกรณีที่อากาศหนาวจัด [6]

วิธีการอบอุ่นร่างกาย British Cycling และ Team Sky [8] ได้ค้นคว้าและทดสอบว่าเป็นการอบอุ่นร่างกายที่ให้ประสิทธิภาพดีที่สุดในนักกีฬาจักรยานลู่อูและถนน จนตอนนี้กลายเป็นมาตรฐานที่หลาย ๆ ทีมทั่วโลกใช้กัน เป้าหมายคือเตรียมความพร้อมให้ร่างกายและกระตุ้นการทำงานของกล้ามเนื้อและระบบต่าง ๆ ด้วยการค่อย ๆ เพิ่มความหนักหน่วงในการออกแรง วิธีการอบอุ่นร่างกายนี้สามารถใช้ได้กับทั้งนักกีฬาอาชีพและมือสมัครเล่นและไม่ได้กำหนดว่าต้องออกแรงกี่วัตต์หรือความเร็วเท่าไร ใช้ความรู้สึก power zone หรืออัตราการเต้นของหัวใจเป็นตัวกำหนดฉะนั้นจึงสามารถใช้กับคนทั่วไปได้

5 นาที : ปั่นเบาๆ รอบขาคนที่

8 นาที : ค่อยๆ เพิ่มความหนักหน่วงของการออกแรงจาก power zone หรือโซนหัวใจเบาสุดไปจนถึงสูงสุด

2 นาที : Recovery ควลงขาจนหายใจเหนื่อย

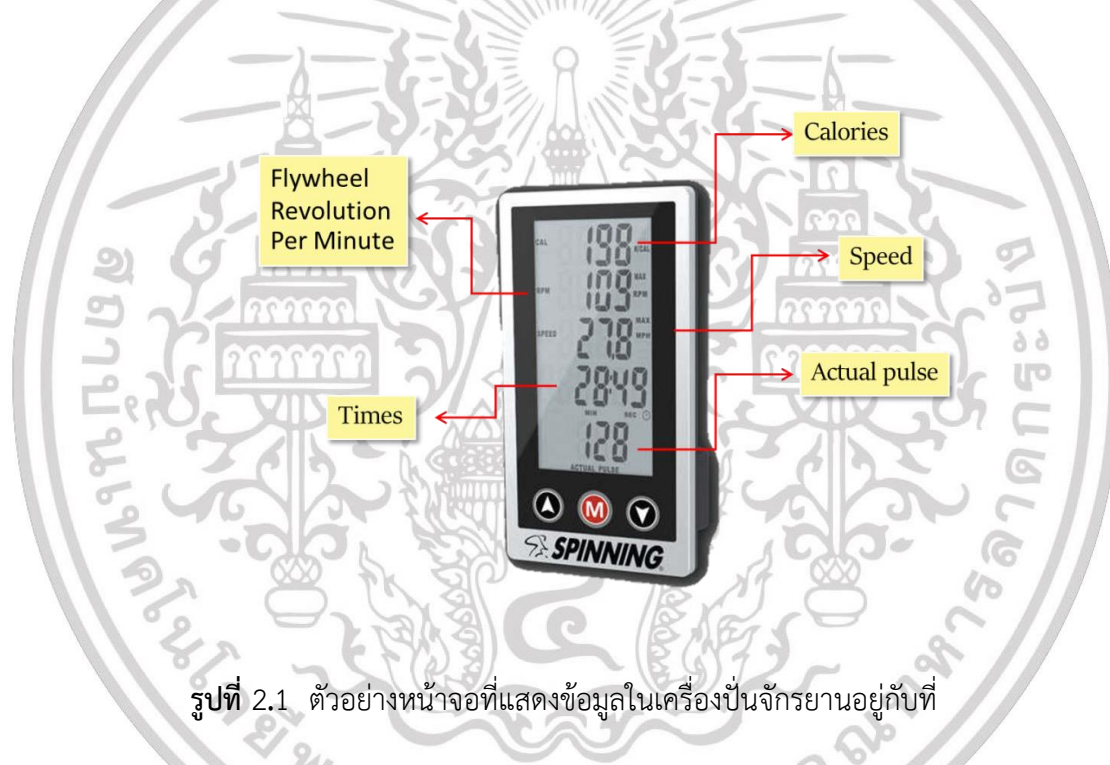
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2 นาที : 3×6 สปริงต์ (สปริงต์สั้นๆ 6 วินาที 3 เซ็ต) เพื่อกระตุ้นการทำงานของกล้ามเนื้อ

3 นาที : Recovery คววงขาอื่นๆ ผ่อนการออกแรงจนหายเหนื่อย

รวมทั้งสิ้น 20 นาที ถ้าอากาศร้อนมากหรือมีเวลาน้อยสามารถลดเวลาในการอบอุ่นร่างกายลงได้ตามสมควร

เครื่องออกกำลังกายอยู่กับที่โดยทั่วไปจะมีหน้าจอแสดงค่าต่างๆ อาทิเช่น ความเร็ว อัตราการเผาผลาญ เวลา รอบการหมุนของล้อต่อนาที ซึ่งพารามิเตอร์เหล่านี้ในงานวิจัยมีแนวคิดที่ต้องการอยากจะทำให้มีโปรแกรมคำนวณระยะทางที่เหมาะสมของแต่ละคน เพื่อเป็นทางเลือกเบื้องต้นสำหรับคนที่อยากรู้ขีดจำกัดการออกกำลังกายของตนเองและเพื่อฟื้นฟูร่างกายอย่างมีประสิทธิภาพไม่ให้ส่งผลเสียภายหลัง



รูปที่ 2.1 ตัวอย่างหน้าจอที่แสดงข้อมูลในเครื่องปั่นจักรยานอยู่กับที่

2.2 โครงข่ายประสาทเทียม

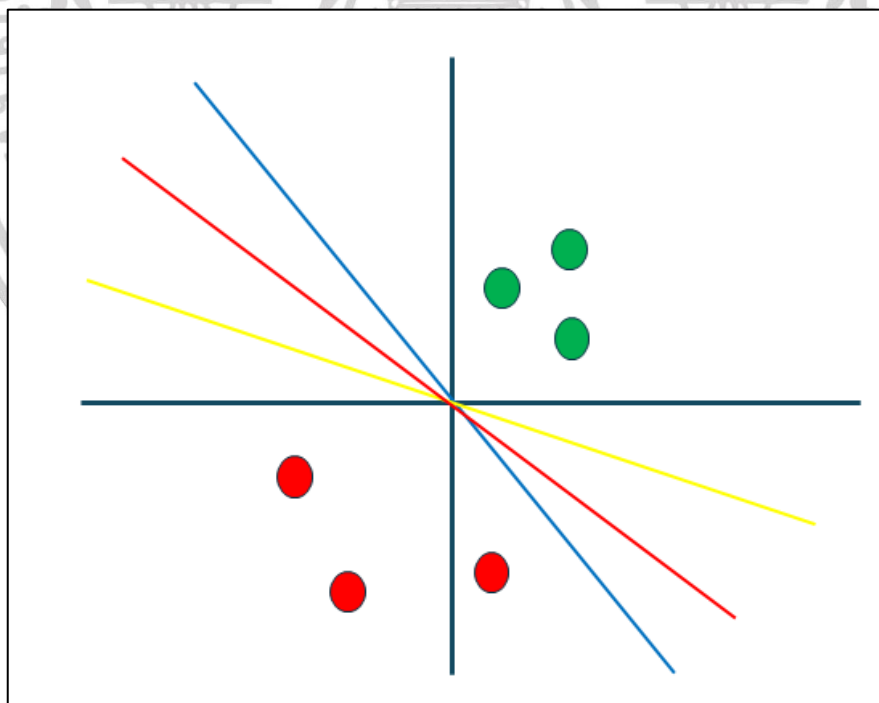
โครงข่ายประสาทเทียมสามารถนำมาประยุกต์ใช้ช่วยแก้ปัญหาได้หลากหลายรูปแบบ เนื่องจากโครงข่ายประสาทเทียมมีความน่าสนใจเพราะระบบโครงข่ายสามารถเรียนรู้ได้เองจากข้อมูลตัวอย่าง และสามารถนำไปประยุกต์ใช้แก้ปัญหาที่มีข้อมูลจำนวนมากและไม่สามารถหาความสัมพันธ์จากข้อมูลได้โดยง่าย การนำโครงข่ายประสาทเทียมมาแก้ปัญหาแต่ละแบบต้องมีการออกแบบโครงข่ายประสาทเทียมขึ้นมาโดยเฉพาะ ดังนั้นการออกแบบโครงข่ายประสาทเทียมที่เหมาะสมกับปัญหาหลาย ๆ วิธีจึงถูกวิจัยและนำเสนอขึ้น ซึ่งแต่ละวิธีต่างมีข้อดีข้อเสียแตกต่างกันไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2.1 วิธีซัพพอร์ตเวกเตอร์แมชชีน (SVM)

ซัพพอร์ตเวกเตอร์แมชชีน (SVM) ถูกเสนอโดย Vapnik [9] เป็นวิธีในการตัดแยกกลุ่มเพื่อจัดประเภท หรือจำแนกประเภทข้อมูลที่มีการนำมาใช้ในด้านการศึกษาประมวลผลภาพโดยอาศัยหลักการของการหา สัมประสิทธิ์ของสมการเพื่อสร้างเส้นแบ่งแยกกลุ่มข้อมูลที่ถูกต้องเข้าสู่กระบวนการสอนให้ระบบเรียนรู้ เป็นวิธีการจำแนกกลุ่มข้อมูล ที่อาศัยระนาบการตัดสินใจที่เรียกว่า ระนาบเกินหรือไฮเปอร์เพลน (Hyperplane) มาใช้ในการจำแนกกลุ่มข้อมูลโดยใช้สมการเส้นตรงในการแบ่งข้อมูลออกเป็น 2 กลุ่มแยกออกจากกัน [10] ซัพพอร์ตเวกเตอร์แมชชีนมีรูปแบบในการเรียนรู้เป็นกระบวนการเลือกแบบจำลองที่เหมาะสมที่สุดจะทำให้ได้ค่าที่เหมาะสมที่สุดเป็นคำตอบ ดังนั้น ซัพพอร์ตเวกเตอร์แมชชีนจึงเป็นที่นิยมและเริ่มนำไปใช้ในงานด้านการรู้จำรูปแบบซึ่งจะเลือกใช้ซัพพอร์ตเวกเตอร์แมชชีนแบบแบ่งกลุ่ม

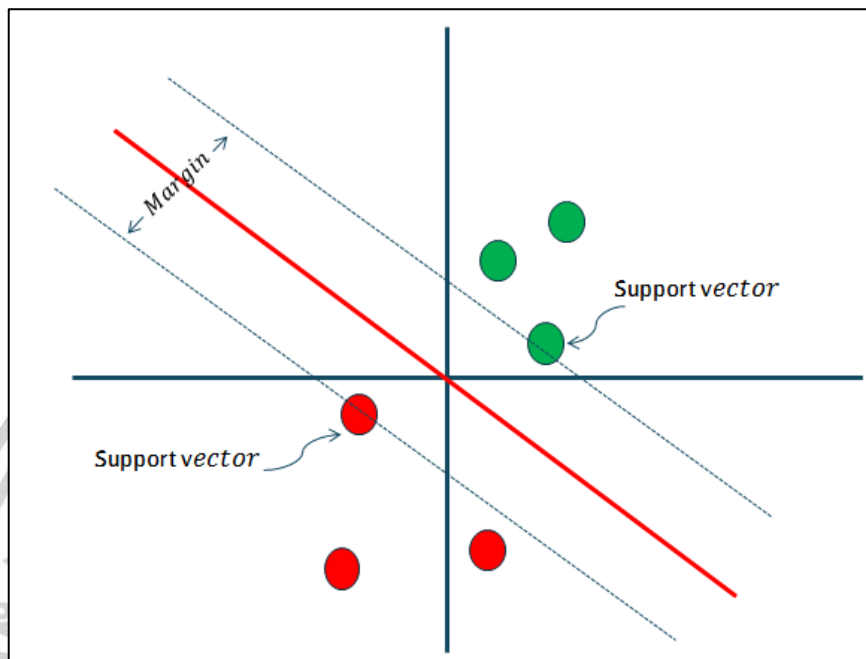
ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการแบ่งกลุ่มข้อมูล จะใช้ระนาบเกินที่เหมาะสมที่สุด (Optimal Hyperplane) ในการแบ่งกลุ่ม การสร้างระนาบเกินที่ใช้ในการแบ่งกลุ่มข้อมูลสามารถสร้างได้หลายแบบ แต่จะมีระนาบเกินที่เหมาะสมที่สุดเพียงระนาบเดียว ที่สามารถรักษาระยะห่างมากที่สุดระหว่างข้อมูล 2 กลุ่มที่ใกล้กัน มากที่สุดได้



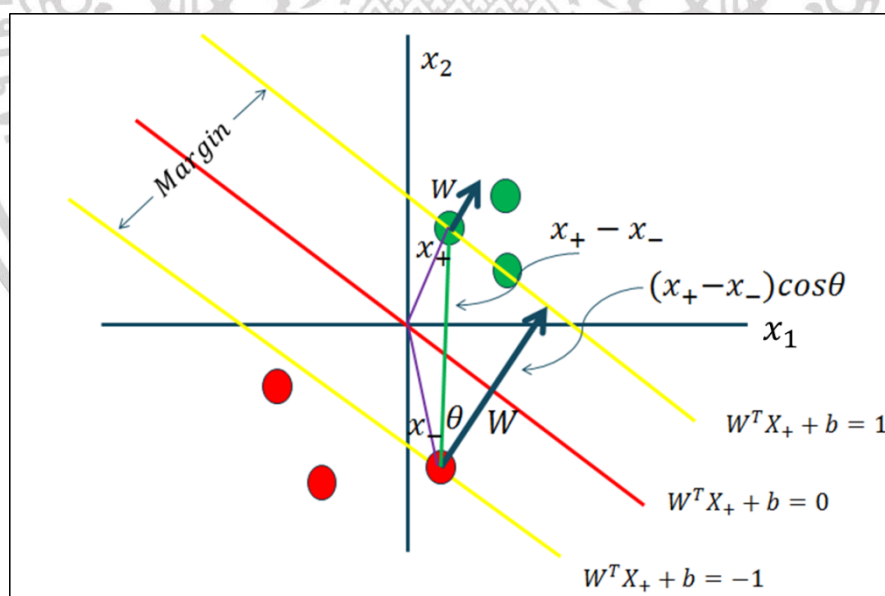
รูปที่ 2.2 แสดงตัวอย่างของเส้นระนาบเกินในการแบ่งข้อมูลออกเป็น 2 คลาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยให้มีระยะขอบ (Margin) ที่มากที่สุดที่แบ่งข้อมูลระหว่าง 2 คลาส และค่าที่อยู่บริเวณเส้นขอบของแต่ละคลาสจะถูกเรียกว่า เวกเตอร์ซัพพอร์ต (Support vector) ดังรูปที่ 2.3



รูปที่ 2.3 แสดงเส้นระยะขอบและเวกเตอร์ซัพพอร์ต



รูปที่ 2.4 แสดงสมการระนาบเกิน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นสมการระนาบเกิน คือ

$$w^T x_+ + b = 1 \quad (2.1)$$

$$w^T x_- + b = -1 \quad (2.2)$$

เมื่อนำสมการที่ (2.1) ลบสมการที่ (2.2)

$$w^T x_+ - w^T x_- = 2 \quad (2.3)$$

$$w^T (x_+ - x_-) = 2 \quad (2.4)$$

เวกเตอร์ที่มีทิศเดียวกับเวกเตอร์ปกติ W และมีขนาดเท่ากับระยะขอบ สามารถหาได้จาก

$$(x_+ - x_-) \cos \theta = (x_+ - x_-) \left(\frac{w^T (x_+ - x_-)}{\|w^T\| \|x_+ - x_-\|} \right) \quad (2.5)$$

แทนสมการที่ (2.4) ในสมการที่ (2.5)

$$\begin{aligned} (x_+ - x_-) \left(\frac{w^T (x_+ - x_-)}{\|w^T\| \|x_+ - x_-\|} \right) &= (x_+ - x_-) \left(\frac{2}{\|w^T\| \|x_+ - x_-\|} \right) \\ &= \frac{2}{\|w\|} \left(\frac{(x_+ - x_-)}{\|x_+ - x_-\|} \right) \end{aligned} \quad (2.6)$$

โดย $\left(\frac{(x_+ - x_-)}{\|x_+ - x_-\|} \right)$ คือเวกเตอร์หนึ่งหน่วยที่มีทิศเดียวกับ w จะได้ค่า Margin = $\frac{2}{\|w\|}$

ดังนั้นปัญหาการหาค่าเหมาะสมของระยะขอบสามารถหาได้จาก

$$\text{Maximize } \frac{2}{\|w\|} \quad \text{โดย } y_i(w^T x_i + b) \geq 1 \quad \text{ทุกค่า } i \quad (2.7)$$

จากปัญหาการหาค่าขอบที่กว้างที่สุดจากปัญหาการหาค่าสูงสุดสามารถเปลี่ยนเป็นปัญหาการหาค่าต่ำสุดได้ดังนี้

$$\frac{2}{\|w\|} = \frac{2}{\sqrt{w^T w}} \propto \frac{2}{w^T w} \quad (2.8)$$

$$\text{Minimize } \frac{1}{2} w^T w \quad \text{โดย } y_i(w^T x_i + b) \geq 1 \quad \text{ทุกค่า } i \quad (2.9)$$

บางปัญหาเราไม่สามารถที่จะหาค่า w ได้โดยตรง ดังนั้นจะแทนด้วยการใช้เงื่อนไข Karush-Kuhn-Tucker โดยใช้วิธีตัวคูณลากรองจ์ซึ่งเป็นวิธีที่เปลี่ยนข้อจำกัดให้อยู่ในรูปแบบที่สามารถใช้

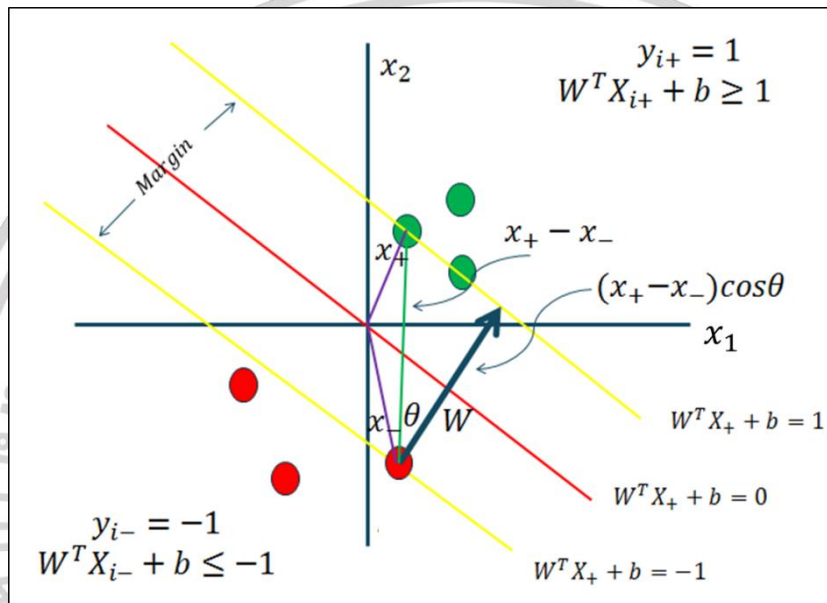
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีการหา จุดสูงสุดและต่ำสุดเหมือนกับกรณีไม่มีข้อจำกัด นั่นคือ การเปลี่ยนให้เป็น Lagrangian function ของปัญหาเดิมสามารถเขียนได้ดังนี้

$$L(w, b, \alpha) = \frac{1}{2} w^T w - \sum_i \alpha_i [y_i (w^T x_i + b) - 1] \quad \text{โดย } \alpha_i [y_i (w^T x_i + b) - 1] = 0$$

และ $\alpha_i \geq 0$ ทุกค่า i (2.10)

เมื่อ α_i คือ Lagrange Multiplier



รูปที่ 2.5 แสดงการแบ่งคลาสที่ $y_{i+} = 1$ และ $y_{i-} = -1$

ดังนั้น

$$\frac{\partial L(w, b, \alpha)}{\partial w} = w - \sum_i y_i \alpha_i x_i = 0 \tag{2.11}$$

$$w = \sum_i y_i \alpha_i x_i \tag{2.12}$$

$$\frac{\partial L(w, b, \alpha)}{\partial b} = \sum_i y_i \alpha_i = 0$$

แทนสมการที่ (2.11) และ (2.12) ในสมการที่ (2.10)

$$\begin{aligned} L(w, b, \alpha) &= \frac{1}{2} w^T w - \sum_i \alpha_i [y_i (w^T x_i + b) - 1] \\ &= \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i^T x_j - \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i^T x_j + \sum_i \alpha_i \end{aligned}$$

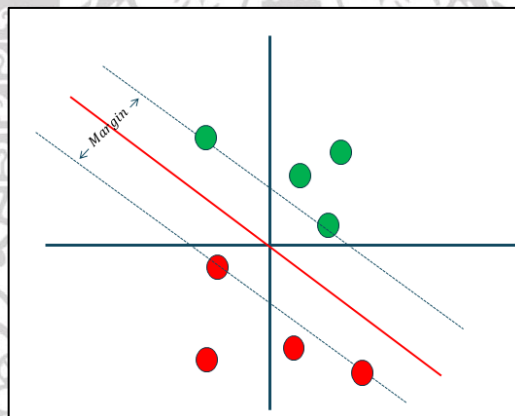
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$= \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i^T x_j \quad (2.13)$$

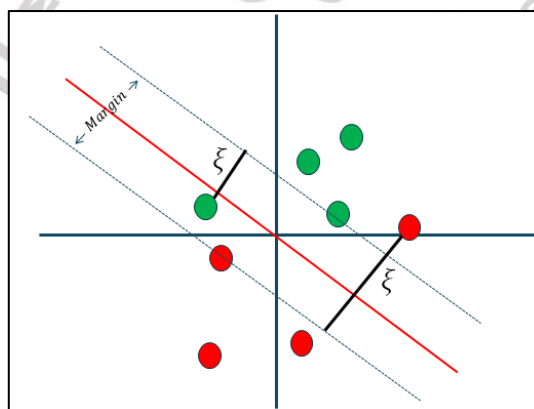
ซึ่งจะได้ปัญหาควบคู่ของปัญหาเดิมดังนี้

$$\text{Maximize } \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i^T x_j \quad \text{โดย } \sum_i y_i \alpha_i = 0, \alpha_i \leq 0 \quad \text{ทุกค่า } i \quad (2.14)$$

จากที่กล่าวมาข้างต้นปัญหาการแบ่งคลาสที่สามารถแบ่งข้อมูลได้ด้วยเส้นตรง (Hard Margin) คือการที่ไม่มีข้อมูลอยู่ในบริเวณขอบเขตของมาร์จินแต่สำหรับบางปัญหาไม่สามารถแบ่งได้ด้วยเส้นตรง ดังนั้นจะแก้ปัญหาดังกล่าวด้วยวิธีการแก้ปัญหามาร์จินอ่อน (Soft Margin) [11] โดยปัญหามาร์จินอ่อนคือการที่อนุญาตให้จุดข้อมูลบางจุดสามารถอยู่ในด้านที่ไม่ถูกต้องได้ ดังนั้นจึงต้องมีการกำหนดตัวแปรสำหรับยอมรับค่าความผิดพลาดโดยการเพิ่มตัวแปร Slack variable (ξ) ดังสมการที่ (2.15) และ (2.16) ดังนี้



รูปที่ 2.6 แสดงปัญหาขอบแข็ง (Hard margin)



รูปที่ 2.7 แสดงปัญหาขอบอ่อน (Soft margin)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

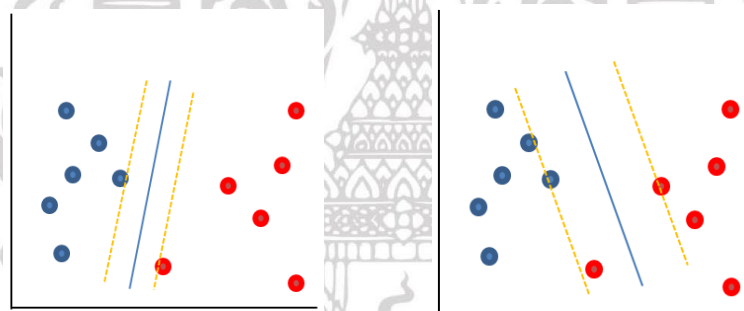
$$w^T x_i + b \geq y - \xi_i \quad \text{เมื่อกำหนดให้ } y=1 \quad (2.15)$$

$$w^T x_i + b \leq y + \xi_i \quad \text{เมื่อกำหนดให้ } y=-1 \quad (2.16)$$

จากการกำหนดค่า $\xi_i > 0$ ทำให้โครงสร้างของซัพพอร์ตเวกเตอร์แมชชีนสามารถเพิ่มระยะแบ่งแยกให้มากที่สุดและลดข้อผิดพลาดในการทำนายให้ต่ำที่สุด ดังสมการที่ (2.17)

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad \text{โดย } y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad \text{สำหรับ } i=1,2,\dots,N \quad (2.17)$$

โดยค่า C คือค่าคงที่ซึ่งเป็นพารามิเตอร์สำหรับกำหนดปริมาณความผิดพลาด ถ้าค่า c มีค่ามากจะทำให้เส้นขอบเขตแคบมากไปด้วยนั่นคือมีจุดข้อมูลที่อยู่ด้านผิดน้อยหรือยอมให้ความผิดพลาดเกิดขึ้นได้น้อยเมื่อนำมาใช้จริงจะทำให้เกิด Overfitting (จดจำข้อมูลเดิมมากเกินไปจนมีผลให้การทำนายผิดพลาด) ถ้าค่า C มีค่าน้อยจะทำให้เส้นขอบเขตมีความกว้างนั่นคือยอมให้มีจุดข้อมูลอยู่ด้านผิดได้หรือยอมให้เกิดความผิดพลาดได้มากจะหลีกเลี่ยงปัญหา Overfitting แต่ไม่ควรมีค่าน้อยจนเกินไป



ค่า c มาก

ค่า c น้อย

รูปที่ 2.8 แสดงตัวอย่างการเปรียบเทียบระหว่างการกำหนดค่า C มากและค่า C น้อย

จะได้ปัญหาควบคู่ของปัญหาการหาค่าต่ำสุดคือ

$$\text{Maximize } \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j x_i^T x_j \quad \text{โดย } \sum_i y_i \alpha_i = 0, 0 \leq \alpha_i \leq c \quad \text{ทุกค่า } i \quad (2.18)$$

จะเห็นว่าฟังก์ชันวัตถุประสงค์คล้ายกับปัญหาเดิมแต่แตกต่างกันตรงเงื่อนไขของค่า α_i ต้องมีค่าน้อยกว่าหรือเท่ากับ C

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากที่กล่าวมาแสดงถึงปริภูมิที่สามารถแบ่งจุดข้อมูลได้ด้วยเส้นตรงหรือเชิงเส้นแต่ในความเป็นจริงอาจไม่สามารถแบ่งจุดข้อมูลในปริภูมินั้น ๆ ได้ด้วยเส้นตรง เราจึงทำการแก้ปัญหาโดยการแปลงปริภูมิไม่เชิงเส้นให้เป็นเชิงเส้น โดยใช้ Kernel Trick

Kernel Trick คือการแปลงปริภูมิให้มีมิติที่สูงขึ้นเพื่อแก้ปัญหาข้อมูลไม่เชิงเส้น [12] ซึ่งฟังก์ชัน Kernel มีหลายแบบดังนี้

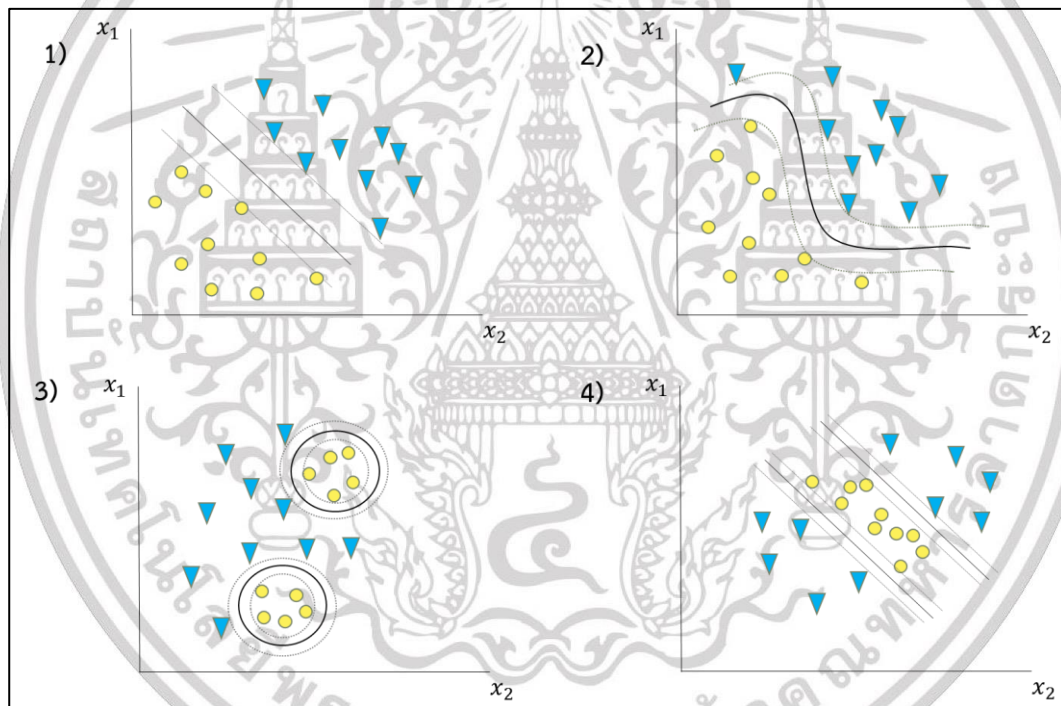
Linear: $K(x_i, y_i) = x_i^T y_i$

Polynomial: $K(x_i, y_i) = (x_i^T y_i + 1)^d$

Radial Basis Function (RBF): $K(x_i, y_i) = e^{-\gamma \|x_i - y_i\|^2}, \gamma > 0$

Sigmoid: $K(x_i, y_i) = \tanh(a \cdot x_i^T y_i + b), \exists [a > 0 \wedge b < 0]$

โดย d, γ, a และ b เป็นพารามิเตอร์ของฟังก์ชันเคอร์เนลมีค่าคงที่ซึ่งขึ้นอยู่กับความเหมาะสม



รูปที่ 2.9 แสดงรูปแบบการแบ่งกลุ่มด้วย Kernel ต่างๆ โดย 1) Linear Kernel 2) Polynomial Kernel 3) Radial Basis Function (RBF) Kernel 4) Sigmoid Kernel

2.2.2 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (SVR)

ในการศึกษาวิธีซัพพอร์ตเวกเตอร์แมชชีนเราใช้เส้นไฮเปอร์เพลนเพื่อแบ่งจุดข้อมูลออกจากกัน K กลุ่ม แต่ในกรณีของวิธีซัพพอร์ตเวกเตอร์รีเกรสชันเราจะนำมาวิเคราะห์ความถดถอยระหว่างอินพุตเวกเตอร์และตัวแปรเอาต์พุตแทนซึ่งเหมาะกับการนำมาแก้ไขปัญหา[13] ซัพพอร์ตเวกเตอร์แมชชีนเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขึ้นสำหรับการถดถอยมีเป้าหมายคือต้องการค้นหาความสัมพันธ์เชิงเส้นระหว่างอินพุตเวกเตอร์ในมิติขนาด n ($x \in R^n$) และตัวแปรเอาต์พุต ($y \in R$) ดังนั้นสมการความถดถอยของวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย จึงคล้ายกับสมการไฮเปอร์เพลนของวิธีซัพพอร์ตเวกเตอร์แมชชีน [14] โดยจะกำหนดระยะเส้นขอบที่ครอบคลุมจุดข้อมูลมากที่สุด แทนด้วย ε โดยเราต้องการประมาณค่าความผิดพลาดของจุดข้อมูลที่อยู่นอกขอบเขต

ให้ข้อมูลฝึกฝนประกอบไปด้วยเซต $\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \subset X \times R$ โดยที่ X คือขนาดของชนิดของข้อมูลนำเข้า ซึ่งในวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย เป้าหมายคือต้องการที่จะหาฟังก์ชัน $f(x)$ จากค่า ε ที่เหมาะสมซึ่งค่าความผิดพลาดไม่ควรมากกว่าค่า ε ซึ่งในกรณีที่เส้นฟังก์ชันเชิงเส้นจะได้รูปแบบดังนี้ [12]

$$F(x) = \langle w, x \rangle + b \quad (2.19)$$

$w \in X, b \in R$ และ สัญลักษณ์ $\langle \cdot, \cdot \rangle$ คือการหาผลคูณของจุดใน x ซึ่งต้องการหาค่าน้อยสุดของ $\|w\|^2$ ดังนั้นจะสามารถเขียนปัญหาการหาค่าเหมาะสมที่สุดจาก

$$\text{Minimize } \frac{1}{2} \|w\|^2 \text{ โดย } y_i - \langle w, x_i \rangle - b \leq \varepsilon \text{ และ } \langle w, x_i \rangle + b - y_i \leq \varepsilon \quad (2.20)$$

จากสมการ (2.20) เป็นการประมาณค่าฟังก์ชัน f จากคู่อันดับ (x_i, y_i) กับค่าคงที่ ε ในกรณีที่เส้นขอบเขตไม่สามารถครอบคลุมจุดข้อมูลทั้งหมดได้ จะทำการเพิ่มตัวแปร $Slack(\xi)$ เพื่อรองรับข้อมูลที่ผิดพลาด ดังนั้นจะได้ปัญหาการหาค่าต่ำสุดดังนี้

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \text{ โดย } y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i, \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ & \text{และ } \xi_i, \xi_i^* \geq 0, i=1, \dots, n \end{aligned} \quad (2.21)$$

โดยที่ $C > 0$ จากปัญหาการหาค่าเหมาะสมในสมการ (2.21) ทำการแก้ปัญหานี้ด้วยการใช้ *Lagrange Multipliers* ด้วยเงื่อนไข *KKT* ได้ดังนี้

$$\begin{aligned} L(w, b, \xi_i, \xi_i^*) = & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) \end{aligned} \quad (2.22)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากปัญหาควคู่สมการ (2.22) เงื่อนไขคือ $\alpha_i, \alpha_i^*, \xi_i, \xi_i^* \geq 0$ เมื่อทำการหาอนุพันธ์ย่อยของ L เทียบกับตัวแปร (w, b, ξ_i^*) จะได้ค่าดังนี้

$$\frac{\partial L}{\partial b} = \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0 \quad (2.23)$$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \quad (2.24)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \quad (2.25)$$

จากสมการ (2.23)-(2.25) นำไปแทนค่าในสมการ (2.22) จะได้ปัญหาควคู่ใหม่ดังนี้

$$\begin{aligned} \text{Maximize} \quad & -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \\ \text{โดย} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (2.26)$$

จะเห็นว่าตัวแปรควคู่ (η_i, η_i^*) ไม่ปรากฏในฟังก์ชันวัตถุประสงค์ของปัญหาควคู่ จากสมการที่ (2.24) สามารถเขียนได้ดังนี้

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i \quad (2.27)$$

ดังนั้น

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (2.28)$$

ในกรณีที่ความสัมพันธ์ของข้อมูลนั้นมีความซับซ้อนจนไม่สามารถแบ่งด้วยเส้นตรงได้จะแก้ปัญหโดยใช้ฟังก์ชัน Kernel $K(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ ที่ $\Phi: X \rightarrow F$ เมื่อ F คือปริภูมิลักษณะเด่น ดังนั้นเราสามารถสร้างสมการซัพพอร์ตเวกเตอร์รีเกรสชันเพื่อทำนายเอาต์พุตจากอินพุตเวกเตอร์ได้ดังสมการที่ (2.29)

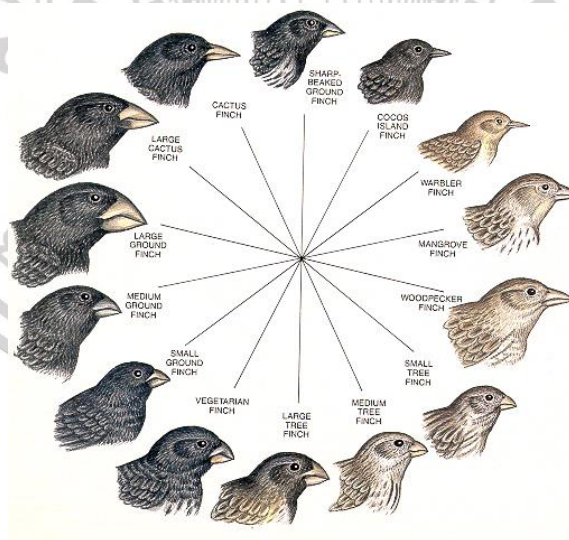
$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (2.29)$$

เมื่อ α_i และ α_i^* คือ ตัวคูณลากรองจ์ $K(x_i, x)$ คือเคอร์เนลฟังก์ชันที่ใช้ในการแปลงข้อมูลให้มีมิติที่สูงขึ้น

2.3 ขั้นตอนวิธีเชิงพันธุกรรม (GA)

แนวคิดของชาร์ลส์ ดาร์วิน (นักธรรมชาติวิทยาชาวอังกฤษ) ได้มาจากข้อมูลทางธรรมชาติที่เขาเก็บรวบรวมได้ ขณะที่เขาเดินทางไปกับเรือสำรวจตามแนวฝั่งทวีปรอบโลก การสำรวจตามแนวฝั่งของทวีปอเมริกาใต้ และหมู่เกาะในมหาสมุทรแปซิฟิก ดาร์วินได้พบว่าพืชและสัตว์ บนพื้นทวีป และหมู่เกาะมีความคล้ายคลึงกันแต่ไม่เหมือนกัน ดาร์วินสังเกตพบว่านกฟินช์ 13 สปีชีส์ พบบนหมู่เกาะกาลาปากอสอยู่ในมหาสมุทรแปซิฟิก มีจงอยปากที่มีขนาดและรูปร่างแตกต่างกันตามความเหมาะสมของประเภทอาหารที่กินแต่ละชนิดกิน จนในปี 1859 ดาร์วิน ตีพิมพ์หนังสือกำเนิดความหลากหลายของสิ่งมีชีวิตโดยการคัดเลือกโดยธรรมชาติ (On the Origin of Species) ซึ่งอธิบายเกี่ยวกับวิวัฒนาการของสิ่งมีชีวิตโดยผ่านกระบวนการคัดเลือกทางธรรมชาติ โดยมีแนวคิดที่ว่า “ตามสภาพธรรมชาติสิ่งมีชีวิตที่มีความสามารถปรับตัวเข้ากับสภาพแวดล้อมได้ดีกว่าพวกอื่น จะสามารถดำรงชีวิตอยู่ และถ่ายทอดลักษณะที่เหมาะสมต่อไป”

ขั้นตอนวิธีเชิงพันธุกรรม (Genetic algorithm หรือ GA) เป็นวิธีที่เลียนแบบวิวัฒนาการทางธรรมชาติของชาร์ลส์ ดาร์วิน ซึ่งแนวคิดนี้ถูกนำมาประยุกต์ใช้โดย Holland (1975) และถูกพัฒนาต่อโดย โกลด์เบิร์ก (Goldberg, 1989) เนื่องจากปัญหาที่พบส่วนใหญ่เป็นปัญหาที่ไม่เที่ยงตรง ซึ่งถ้าหากต้องการคำตอบที่เที่ยงตรงและมีความแม่นยำสูงมากก็ย่อมมีค่าใช้จ่ายที่สูงขึ้น วิธีการแก้ปัญหาคือการหาคำตอบที่ใกล้เคียงที่สามารถยอมรับได้ วิธีค้นหาคำตอบที่ได้อย่างหนึ่งได้แก่วิธีเชิงพันธุกรรมซึ่งจัดอยู่ในประเภทวิธีการฮิวริสติก



รูปที่ 2.10 นกฟินช์ (Finch) ตามภูมิภาคต่างๆ (อ้างอิง : www.sites.google.com)

เนื่องจากขั้นตอนวิธีเชิงพันธุกรรมมีรากฐานมาจากทฤษฎีวิวัฒนาการจึงมีศัพท์เฉพาะทางชีววิทยาที่สำคัญเพื่อนำมาแก้ปัญหาไว้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.2 ศัพท์เฉพาะในวิธีเชิงพันธุกรรม

คำศัพท์	หมายถึง
ยีน (Gene)	ค่าที่เป็นไปได้ในแต่ละหน่วยย่อยของโครโมโซม
โครโมโซม (Chromosome)	กลุ่มผลลัพธ์ที่เป็นไปได้ของปัญหา
ประชากร (Population)	กลุ่มของโครโมโซม
การสืบพันธุ์แบบใช้เพศ	การที่ประชากรรุ่นต่อไป(offspring)เกิดขึ้นจากรุ่นก่อนมากกว่า 1 ตัวเช่น ไขว้เปลี่ยน(Crossover)
การสืบพันธุ์แบบไม่ใช้เพศ	การที่ประชากรรุ่นต่อไป (offspring) เกิดขึ้นจากรุ่นก่อนเพียงตัวเดียวเช่น การกลายพันธุ์ (Mutation)
การคัดเลือก	การอาศัยค่า Fitness หรือค่าความเหมาะสมในการคัดเลือกรุ่นถัดไป

2.3.1 กระบวนการขั้นตอนวิธีเชิงพันธุกรรม

1) ค่าประชากรเริ่มต้น (Initial Population) เริ่มจากการสุ่มคำตอบซึ่งเป็นคำตอบที่อยู่ในขอบเขตของคำตอบขึ้นมาตามจำนวนของประชากรที่กำหนดไว้แล้วหลังจากนั้นทำการเข้ารหัสชุดคำตอบนั้นจะเรียกว่าโครโมโซมซึ่งโครโมโซมประกอบไปด้วยยีนที่เรียงต่อกัน ซึ่งการเข้ารหัสทางพันธุกรรมสามารถทำได้หลายวิธีเช่น

การแทนด้วยเลขฐานสอง (Binary Encoding) เป็นการออกแบบโครโมโซมที่ใช้ตัวเลข 1 และ 0 ซึ่งนิยมใช้กับผลเฉลยที่แต่ละทางเลือกเป็นใช่หรือไม่ใช่ตัวอย่างการแทนรหัสเช่น 00110101011

การแทนด้วยจำนวนจริง (Real Number Encoding) ใช้กับผลเฉลยที่เกี่ยวข้องกับข้อมูลเชิงปริมาณหรือออกแบบโครโมโซมโดยใช้อักษรเป็นตัวแทนของค่าคำตอบ

การแทนด้วยการเรียงสับเปลี่ยน (Permutation Encoding) เป็นการออกแบบโครโมโซมที่ใช้เป็นตัวเลขทั่วไปซึ่งนิยมใช้กับข้อมูลที่ลำดับก่อนหลังมีผลกับผลเฉลย เช่น ปัญหาการจัดลำดับการผลิต

การแทนด้วยรายการของกฎ (Rule Encoding) เช่น การตัดสินใจหรือการวินิจฉัย

2) ฟังก์ชันความเหมาะสม (Fitness Function) ในขั้นตอนนี้โครโมโซมทั้งหมดจะถูกประเมินค่าความเหมาะสมของโอกาสในการอยู่รอดซึ่งจะใช้สมการเป้าหมายในการวัดค่าความน่าจะเป็น ซึ่งสมการเป้าหมายจะขึ้นอยู่กับแต่ละปัญหา

3) พ่อแม่ (Parent) ในขั้นตอนนี้จะทำการคัดเลือกโครโมโซมโดยโครโมโซมที่มีค่าความเหมาะสมสูงจะถูกคัดเลือก (Selection) ให้เป็นพ่อ-แม่ ซึ่งการคัดเลือกมีหลายวิธีเช่น วิธีแบ่งเป็นสัดส่วน (Proportionate Selection) วิธีของโบลต์ซมันน์ (Boltzmann Selection) วิธีการจัดอันดับ

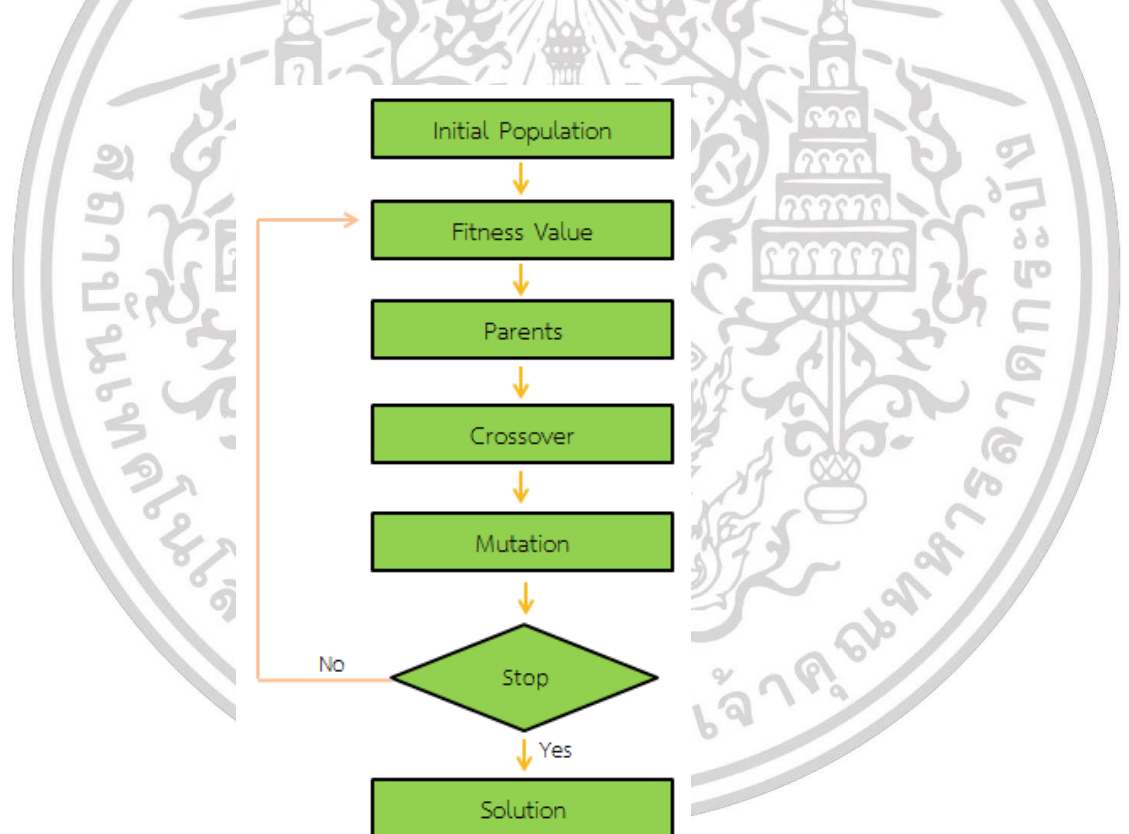
(Ranking Selection) วิธีการแข่งขัน (Tournament Selection) และ วิธีการชักตัวอย่างแบบวงล้อ

ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เล็ต (Roulette Wheel Sampling) จากที่กล่าวมาข้างต้นวิธีที่นิยมที่สุดคือ วิธีการชักตัวอย่างแบบวงล้อเล็ต

4) การสลับสายพันธุ์ (Crossover) เป็นกระบวนการสืบพันธุ์แบบอาศัยเพศ ซึ่งจะทำให้การสลับยีนในโครโมโซมโดยจะกำหนดความน่าจะเป็นของการสลับสายพันธุ์ให้มิต่ำค่ามาก ซึ่งกระบวนการในการสลับสายพันธุ์จะเกิดขึ้นก็ต่อเมื่อค่าการสุ่มในการสลับสายพันธุ์มีค่าน้อยกว่าค่าความน่าจะเป็นในการสลับสายพันธุ์ เช่น กำหนดความน่าจะเป็นในการสลับสายพันธุ์คือ 0.7 ถ้าค่าสุ่มในการสลับสายพันธุ์คือ 0.65 จะเกิดการสลับสายพันธุ์ขึ้น การสลับสายพันธุ์มีหลายรูปแบบเช่น การไขว้เปลี่ยนแบบจุดเดี่ยว(Single-point) การไขว้เปลี่ยนแบบสองจุด (Two-point) และการไขว้เปลี่ยนแบบเอกรูป (Uniform)

5) การกลายพันธุ์ (Mutation) เป็นกระบวนการสืบพันธุ์แบบไม่อาศัยเพศซึ่งโครโมโซมใหม่เกิดจากโครโมโซมรุ่นก่อนโครโมโซมเดียว การกลายพันธุ์จะเกิดขึ้นก็ต่อเมื่อค่าการสุ่มในการกลายพันธุ์มีค่า



รูปที่ 2.11 แสดงขั้นตอนวิธีเชิงพันธุกรรม

น้อยกว่าค่าความน่าจะเป็นในการกลายพันธุ์โดยจะกำหนดความน่าจะเป็นของการกลายพันธุ์ให้มิต่ำค่าน้อยๆ เช่น 0.01 การกลายพันธุ์ทำได้สองรูปแบบคือ การกลายพันธุ์แบบจุดและการกลายพันธุ์แบบสลับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6) เงื่อนไขการหยุดการทำงาน ในขั้นตอนนี้จะเป็นตัวกำหนดการสิ้นสุดการทำงานเช่น กำหนดให้สิ้นสุดการทำงานเมื่อครบ 100 รอบหรือสิ้นสุดการวนรอบเมื่อค่าความเหมาะสม (Fitness Function) ไม่สามารถพัฒนาได้

2.4 การวัดความถูกต้องของการพยากรณ์

การประเมินประสิทธิภาพของแบบจำลองระยะทางการ จะใช้ค่าที่ผิดพลาดที่เกิดขึ้นของการพยากรณ์ชุดข้อมูลทดสอบโดยเปรียบเทียบค่าผลลัพธ์ที่แท้จริงกับค่าที่ได้จากการพยากรณ์โดยมีวิธีวัดค่า 2 วิธี

2.4.1 รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE)

เป็นหนึ่งในวิธีที่ใช้บ่อยที่สุดในการประเมินประสิทธิภาพของการพยากรณ์ซึ่งจะใช้วัดว่าค่าพยากรณ์ห่างจากค่าจริงแค่ไหนโดยใช้ระยะทางแบบยูคลิด (Euclidean distance) ทั่วไปจะใช้รากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ยในภูมิอากาศวิทยา การพยากรณ์ และการวิเคราะห์การถดถอยเพื่อตรวจสอบผลการทดลอง

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (2.31)$$

2.4.2 เปอร์เซนต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE)

เป็นค่าวัดความถูกต้องของการพยากรณ์ที่วัดจากขนาดของค่าความคลาดเคลื่อนของการพยากรณ์เทียบกับค่าจริง โดยไม่คิดเครื่องหมาย จึงเหมาะที่จะใช้กับการเปรียบเทียบวิธีการพยากรณ์เดียวกันหรือการเปรียบเทียบวิธีการพยากรณ์หลายวิธี

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad (2.32)$$

โดยกำหนดให้ตัวแปร y_t คือข้อมูลจริง ตัวแปร \hat{y}_t คือความคลาดเคลื่อนและตัวแปร n คือ จำนวนข้อมูลทั้งหมด

2.5 งานวิจัยที่เกี่ยวข้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

A. Zander, E. Passmore, C. Mason and C. Rissel [15] ทำการศึกษาเชิงคุณภาพสำรวจตัวกระตุ้น สิ่งกระตุ้น และอุปสรรคในการปั่นจักรยานในหมู่ผู้สูงอายุผ่านโครงการส่งเสริมการปั่นจักรยานที่กำหนดเป้าหมายตามอายุ วิธีการคือให้ผู้ใหญ่จำนวน 17 คนที่มีอายุ 50–75 ปีเข้าร่วมโปรแกรมส่งเสริมการปั่นจักรยาน 12 สัปดาห์ ซึ่งรวมถึงหลักสูตรทักษะการปั่นจักรยาน จากรายงานผลลัพธ์การปั่นจักรยาน ได้แก่ คุณภาพชีวิตที่ดีขึ้น (สุขภาพจิตที่ดีขึ้น ผลประโยชน์ทางสังคม และการเสริมสร้างพลังอำนาจ) และสุขภาพร่างกายที่ดีขึ้น

S. Titze, B. D. Geus, P. Krenn, A. Bauman, P. Oja, B. R. Nash and T. Kohlberger [16] ศึกษาเกี่ยวกับความสัมพันธ์ระหว่างการปั่นจักรยานกับสมรรถภาพหัวใจและหลอดเลือด ผลจากการศึกษาในผู้ใหญ่วัยทำงานบ่งชี้ว่าสมรรถภาพของหัวใจและหลอดเลือดดีขึ้นอย่างสม่ำเสมอและปัจจัยเสี่ยงโรคหัวใจและหลอดเลือดบางส่วนดีขึ้นเนื่องจากการปั่นจักรยาน

ในงานวิจัยหลายงานได้นำวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression) ไปประยุกต์ใช้เพื่อแก้ปัญหาการพยากรณ์หรือเพื่อเพิ่มประสิทธิภาพการพยากรณ์ซึ่งสามารถนำไปประยุกต์ได้หลายด้านเช่น ด้านการแพทย์เพื่อวินิจฉัยโรค นอกจากนี้บางงานวิจัยยังนำเสนอวิธีที่จะช่วยเพิ่มประสิทธิภาพการพยากรณ์โดยใช้วิธีเชิงพันธุกรรมเข้ามาประยุกต์ใช้กับวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย

โดยเทคนิคทั่วไปที่ใช้ในการพยากรณ์จะเป็นวิธีการวิเคราะห์การถดถอยเชิงพหุคูณและวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยเป็นการวิเคราะห์ความถดถอยระหว่างอินพุตที่อยู่ในรูปแบบเวกเตอร์กับตัวแปรตามด้วยหลักการด้านการเรียนรู้ของเครื่อง [17] วิธีนี้ให้ความแม่นยำสูง [18-21] แต่ SVR มีข้อจำกัดในการใช้งานเนื่องจากผู้ใช้ต้องกำหนดค่าพารามิเตอร์ต่าง ๆ ให้เหมาะสมเพื่อให้ได้แบบจำลอง SVR ที่มีประสิทธิภาพ ดังนั้นเราจึงเสนอวิธี Genetic Algorithm-Based Support Vector Regression (GA-SVR) โดยการใช้ วิธีเชิงพันธุกรรม (Genetic Algorithm : GA) ในการค้นหาค่าพารามิเตอร์ใน SVR ที่เหมาะสม GA-SVR สามารถนำไปประยุกต์ใช้เพื่อแก้ปัญหาการพยากรณ์หรือเพื่อเพิ่มประสิทธิภาพการพยากรณ์ซึ่งสามารถนำไปประยุกต์ได้หลายด้าน

C. C. Yang and M. D. Shieh [22] งานวิจัยนี้เสนอวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยเพื่อพัฒนาแบบจำลองที่คาดการณ์การตอบสนองทางอารมณ์ของผู้บริโภค (consumers' affective responses : CARs) สำหรับการออกแบบผลิตภัณฑ์ ประการแรกดูค่าที่ใช้อธิบายต่อตัวอย่างผลิตภัณฑ์ ประการที่สองดูคุณลักษณะของผลิตภัณฑ์ (product form features : PFFs) ทำการเก็บข้อมูลการประเมินโดยรวมจากแบบสอบถาม แบบจำลองถูกสร้างขึ้นโดยใช้ SVR และใช้วิธี(RCGA)เพื่อค้นหาค่าพารามิเตอร์ที่เหมาะสมที่สุด จากนั้นทำการเปรียบเทียบประสิทธิภาพการทำนายระหว่าง Support Vector Regression with real-coded genetic algorithm (SVR-เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

RCGA) Support Vector Regression with 5-fold cross-validation (SVR-5FCV) back-propagation neural network with 5-fold cross-validation (BPNN -5FCV) ผลการทดลองโดยใช้ชุดข้อมูลบนโทรศัพท์มือถือและสวิตเตอร์ไฟฟ้าแสดงให้เห็นว่า SVR ทำงานได้ดีกว่า BPNN นอกจากนี้ RCGA สำหรับการเพิ่มประสิทธิภาพพารามิเตอร์ในวิธี SVR ยังสะดวกต่อการใช้งานจริงในการออกแบบรูปแบบผลิตภัณฑ์มากกว่าวิธี cross-validation (CV) ที่ใช้เวลานาน

M. H. Zangoeei, J. Habibi and R. Alizadehsani [23] งานวิจัยนี้อธิบายเกี่ยวกับการวินิจฉัยโรคทั้งโรคที่ไม่ติดต่อเช่นโรคมะเร็งเต้านม โรคตับ โรคเบาหวาน โรคที่ติดต่อได้คือโรคไวรัสตับอักเสบ โดยจะวินิจฉัยการเป็นโรคชนิดต่างๆของผู้ป่วยจากการพยากรณ์ที่ประยุกต์ใช้วิธี SVR ร่วมกับ Non-dominated Sorting Genetic Algorithm II (NSGAI) ผลการวัดค่าความถูกต้องเทียบกับวิธีอื่น ๆ ที่เคยมีการศึกษามา เช่น การพยากรณ์โรคตับด้วยวิธีอื่นๆ อาทิ AWAIIS Naïve Bayes และ BNNF การพยากรณ์โรคมะเร็งเต้านมด้วยวิธีอื่นๆ อาทิ RIAC LDA และ Fuzzy-GA1 การพยากรณ์โรคเบาหวานด้วยวิธีอื่นๆ อาทิ Hybtid system Naïve Bayes และ SVM การพยากรณ์โรคไวรัสตับอักเสบด้วยวิธีอื่นๆ อาทิ BNND Weighted 9NN และ PCA-LSSVM จากวิธีที่นำมาใช้พยากรณ์โรคต่างๆเมื่อนำมาเปรียบเทียบความแม่นยำในการพยากรณ์พบว่า วิธี SVR ร่วมกับ NSGAI มีความแม่นยำในการพยากรณ์สูงกว่าวิธีอื่น ๆ

X. Tang, L. Wang, J. Chang, J. Chen and V. S. Sheng [24] ได้เสนอแบบจำลองแบบผสมผสานที่ไม่เคยมีมาก่อนโดยอิงจากการแยกย่อยข้อมูลที่คลุมเครือ (fuzzy information) ซึ่งผสมผสานร่วมกับแบบจำลองของ GA-SVR และ Autoregressive Integrated Moving Average (ARIMA) ในการทำนายดัชนีราคาผู้ผลิต (Producer Price Index : PPI) จากการวิจัยแสดงให้เห็นว่าค่า PPI ที่คาดการณ์โดยแบบจำลองผสมผสานนี้มีความแม่นยำมากกว่าที่คาดการณ์โดยแบบจำลองอื่น ๆ

M. L. Guo, D. Li, C. Du, Z. Jia, X. Qin, L. Chen, L. Sheng and H. Li [25] การพยากรณ์การจราจรที่หนาแน่นของเมืองซินเจียงในวันหยุดสุดสัปดาห์ จากการศึกษาก่อนหน้านี้มีการใช้วิธี ARIMA และวิธี SVM ในการพยากรณ์การจราจรในช่วงวันหยุดสุดสัปดาห์ ในงานวิจัยนี้จึงนำเสนอวิธี GA-SVR โดยใช้ radial basis function เป็น kernel functions และเปรียบเทียบวิธี SVR แบบดั้งเดิม วิธี GA-SVR และวิธี Back Propagation Neural Network (BP Neural Network) ผลลัพธ์ที่ได้แสดงให้เห็นว่า GA-SVR ให้ความแม่นยำในการพยากรณ์สูงสุดและใช้เวลาน้อยลง ซึ่งเป็นวิธีที่มีประสิทธิภาพในการพยากรณ์สภาพการจราจรที่หนาแน่น

Y. S. Kao, K. Nawata and C. Y. Huang [26] การคาดการณ์การใช้พลังงานของประเทศเนื่องจากข้อมูลมีลักษณะไม่เชิงเส้นจึงทำให้ไม่สามารถคาดการณ์ได้อย่างแม่นยำด้วยวิธีการพยากรณ์แบบเดิม ดังนั้นจึงมีการเสนอกรอบการคาดการณ์ด้วยการผสมผสานหลายวิธี ประกอบด้วยเอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Ensemble Empirical Mode Decomposition (EEMD) Autoregressive Integrated Moving Average (ARIMA) SVR และ GA ได้เป็นวิธี EEMD-ARIMA-GA-SVR จากผลการศึกษา การผสมผสานหลายวิธีที่เสนอนั้นสามารถนำไปใช้ได้ เมื่อเทียบกับผลการคาดการณ์วิธีอื่น ๆ พบว่าให้ความแม่นยำที่ดีกว่า ระบบวิธีผสมผสานที่กล่าวมานี้เป็นพื้นฐานสำหรับการจัดการพลังงานและการกำหนดนโยบาย

S. Liu, H. Tai, Q. Ding, D. Li, L. Xu and Y. Wei [27] แบบจำลองการคาดการณ์คุณภาพน้ำสำหรับการเพาะเลี้ยงสัตว์น้ำ การศึกษานี้แนะนำแนวทางไฮบริดที่เรียกว่า real-value genetic algorithm support vector regression (RGA-SVR) ซึ่งค้นหาพารามิเตอร์ SVR ที่เหมาะสมที่สุด จากนั้นจึงนำพารามิเตอร์ที่เหมาะสมที่สุดมาใช้เพื่อสร้างแบบจำลอง SVR แบบจำลองนี้ใช้เพื่อทำนายข้อมูลคุณภาพน้ำเพาะเลี้ยงสัตว์น้ำที่รวบรวมจากโรงงานสัตว์น้ำของ YiXing ในประเทศจีน ผลการทดลองแสดงให้เห็นว่า RGA-SVR มีประสิทธิภาพเหนือกว่าแบบจำลอง SVR และ back-propagation (BP) แบบดั้งเดิม โดยอิงจากค่าคลาดเคลื่อน RMSE และค่า MAPE แบบจำลอง RGA-SVR นี้ได้รับการพิสูจน์แล้วว่าเป็นแนวทางที่มีประสิทธิภาพในการคาดการณ์คุณภาพน้ำสำหรับการเพาะเลี้ยงสัตว์น้ำ

H. Ahn, S. Kim and J. K. Kim. [28] สร้างแบบจำลองการประเมินสถานะทางอารมณ์ตามการแสดงออกทางสีหน้า การศึกษาส่วนใหญ่ใช้วิธี Multiple Regression Analysis (MRA) Artificial Neural Network (ANN) และ Support Vector Regression (SVR) แต่มีความแม่นยำในการทำนายค่อนข้างต่ำ เพื่อปรับปรุงประสิทธิภาพการพยากรณ์ของแบบจำลองการจึงเสนอแบบจำลอง GA-SVR โดยใช้ GA เพื่อเพิ่มประสิทธิภาพพารามิเตอร์เคอร์เนลของ SVR ซึ่งเคอร์เนลที่เลือกใช้คือ Radial basis Function (RBF) เป็นเคอร์เนลที่นิยมใช้ในหลายๆงานวิจัย [29 , 30] เพื่อตรวจสอบประสิทธิภาพของแบบจำลอง GA-SVR เราเก็บรวบรวมชุดข้อมูลจริงของการตอบสนองทางใบหน้าและสถานะทางอารมณ์ผ่านแบบสำรวจ ใช้แบบจำลอง GA-SVR เปรียบเทียบกับแบบจำลอง MRA, ANN และ SVR กับชุดข้อมูลเดียวกัน พบว่า GA-SVR มีประสิทธิภาพเหนือกว่าวิธีอื่น ๆ

บทที่ 3

วิธีดำเนินการวิจัย

ในบทนี้จะนำทฤษฎีบทต่างๆ มาใช้ในการสร้างแบบจำลองจากที่ได้กล่าวมาในบทก่อนหน้า โดยจะอธิบายตัวแปรที่นำมาใช้ อธิบายวิธีการสร้างแบบจำลองด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมและวิธีอื่น ๆ

3.1 ข้อมูลที่นำมาใช้

การดำเนินงานของการวิเคราะห์การศึกษาเรื่องการพยากรณ์ระยะทางการปั่นจักรยานเพื่อสุขภาพโดยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมโดยผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

- 1.การกำหนดประชากรและกลุ่มตัวอย่าง
- 2.การสร้างเครื่องมือที่ใช้ในการวิเคราะห์
- 3.การเก็บรวบรวมข้อมูล
- 4.การจัดกระทำข้อมูล

ขอบเขตของการวิเคราะห์ประกอบไปด้วยประชากร คือ คนที่ปั่นจักรยานชาวไทยที่อาศัยอยู่ในกรุงเทพมหานครและอยู่ในเขตลาดกระบัง ในการศึกษาวิจัยในครั้งนี้ศึกษาเกี่ยวกับการวิเคราะห์ระยะทางที่เหมาะสมในการปั่นจักรยาน เนื่องจากในเขตลาดกระบังมีทางสำหรับจักรยานอยู่หลายแห่งและเป็นที่ยอมรับของคนปั่นจักรยานชาวไทยที่อาศัยอยู่ในกรุงเทพมหานคร

กลุ่มตัวอย่างที่ใช้ในการวิจัยได้แก่ คนที่ปั่นจักรยานชาวไทยที่อาศัยอยู่ในกรุงเทพมหานครและอยู่ในเขตลาดกระบัง จำนวน 105 ตัวอย่าง การเลือกกลุ่มตัวอย่างในงานวิจัยครั้งนี้ ผู้วิจัยจะเก็บข้อมูลจากผู้ปั่นจักรยานที่เป็นคนไทยที่อยู่ในเขตลาดกระบัง โดยใช้วิธีการสุ่มแบบการเลือกตัวอย่างแบบตามขั้นตอนประกอบด้วย

ขั้นตอนที่ 1 ใช้การเลือกตัวอย่างแบบเจาะจง (Purposive sampling) เพื่อเจาะจงสถานที่ที่จะใช้ในการสอบถามกลุ่มเป้าหมาย ได้แก่ คนที่ปั่นจักรยานที่เป็นชาวไทยที่อาศัยอยู่ในกรุงเทพมหานครและอยู่ในเขตลาดกระบัง

ขั้นตอนที่ 2 ใช้การเลือกตัวอย่างแบบโควตา (Quota sampling) เพื่อกำหนดจำนวนตัวอย่างที่จะเก็บ โดยเก็บจำนวน 105 ชุด

ขั้นตอนที่ 3 ใช้การเลือกตัวอย่างแบบสะดวก (Convenience sampling) คือการเข้าไปแจกแบบสอบถามจากกลุ่มตัวอย่างที่ได้กำหนดไว้แล้วในขั้นตอนที่ 1 และ 2 ให้ครบตามจำนวนแบบสอบถามจำนวน 105 ชุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 การสร้างเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลครั้งนี้ คือ แบบสอบถาม ซึ่งเป็นการรวบรวมข้อมูล จากคนที่ปั่นจักรยานชาวไทยที่อาศัยอยู่ในกรุงเทพมหานครและอยู่ในเขตลาดกระบัง โดยผู้วิจัยสร้างขึ้น โดยการศึกษาค้นคว้าจาก หนังสือและเอกสารต่าง ๆ โดยลักษณะของแบบสอบถามแบ่งออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 แบบสอบถามเกี่ยวกับข้อมูลส่วนบุคคลของคนปั่นจักรยาน ได้แก่

1. เพศ เป็นแบบสอบถามชนิดปลายปิด (Close - ended response question) โดยใช้ระดับการวัดข้อมูลประเภทนามบัญญัติ (Nominal Scale) ได้แก่ เพศชาย และ เพศหญิง

3.1 เพศชาย

3.2 เพศหญิง

2. อายุ เป็นแบบสอบถามชนิดปลายเปิด Open-ended Question โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

3. โรคประจำตัวเป็นแบบสอบถามชนิดปลายปิด (Close - ended response question) โดยใช้ระดับการวัดข้อมูลประเภทนามบัญญัติ (Nominal Scale) ได้แก่

3.1 ไม่มี

3.2 มี โปรดระบุ

4. น้ำหนัก เป็นแบบสอบถามชนิดปลายเปิด Open-ended Question โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

5. ส่วนสูง เป็นแบบสอบถามชนิดปลายเปิด Open-ended Question โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

ส่วนที่ 2 แบบสอบถามเกี่ยวกับการใช้จักรยาน

6. แบบสอบถามเกี่ยวกับประเภทของจักรยานเป็นแบบสอบถามชนิดปลายปิด (Close - ended response question) โดยใช้ระดับการวัดข้อมูลประเภทนามบัญญัติ (Nominal Scale) ได้แก่

6.1 Mountain Bike (จักรยานเสือภูเขา)

6.2 Fixed Gear (จักรยานแข่ง แบบไม่มีเกียร์)

6.3 Road Bike (จักรยานขับบนถนนทางเรียบ)

6.4 Stationary Bike (จักรยานในฟิตเนส)

6.5 BMX (จักรยานวิบาก)

6.6 Utility Bike (จักรยานทั่วไป ปั่นแถวบ้าน ปั่นไปจ่ายตลาด)

6.7 อื่น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. แบบสอบถามเกี่ยวกับอาการของโรคหลังปั่นจักรยานเฉพาะผู้ที่มีโรคประจำตัว เป็นแบบสอบถามชนิดปลายปิด (Close - ended response question) โดยใช้ระดับการวัดข้อมูลประเภทนามบัญญัติ (Nominal Scale) ได้แก่

- 7.1 ดีขึ้น
- 7.2 ทรงตัว
- 7.3 แย่ลง
- 7.4 ไม่ทราบ

8. แบบสอบถามเกี่ยวกับระยะทางในการปั่นจักรยานที่ผู้ทำแบบสอบถามรู้สึกเหนื่อยหรือชืดจำกัดของระยะทางในการปั่นจักรยานซึ่งมีหน่วยเป็นกิโลเมตร เป็นแบบสอบถามชนิดปลายเปิด Open-ended Question โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

9. แบบสอบถามเกี่ยวกับเวลาในการปั่นจักรยานที่ผู้ทำแบบสอบถามรู้สึกเหนื่อยหรือชืดจำกัดของเวลาในการปั่นจักรยานซึ่งมีหน่วยเป็นนาที เป็นแบบสอบถามชนิดปลายเปิด Open-ended Question โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

10. แบบสอบถามเกี่ยวกับอัตราการเต้นของหัวใจสูงสุดขณะเหนื่อย ซึ่งมีหน่วยเป็นครั้งต่อนาทีเป็นแบบสอบถามชนิดปลายเปิด (Open-ended Question) โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

11. แบบสอบถามเกี่ยวกับพฤติกรรมกรรมการปั่นจักรยานของผู้ทำแบบสอบถามว่าในหนึ่งสัปดาห์ปั่นจักรยานกี่ครั้ง ซึ่งเป็นแบบสอบถามชนิดปลายเปิด (Open-ended Question) โดยใช้ระดับมาตราอัตราส่วน (Ratio scale)

วิธีการหาคุณภาพเครื่องมือ

1. นำแบบสอบถามที่ได้ทำการสร้างขึ้นให้อาจารย์ที่ปรึกษาพิจารณาความถูกต้อง และให้ข้อเสนอแนะเพิ่มเติม
2. แก้ไขปรับปรุงแบบสอบถามตามคำแนะนำของอาจารย์ที่ปรึกษา
3. นำแบบสอบถามที่ปรับปรุงแล้วให้ผู้เชี่ยวชาญตรวจสอบความเที่ยงตรง
4. แก้ไขปรับปรุงแบบสอบถามตามคำแนะนำของผู้เชี่ยวชาญ จากนั้นนำเสนออาจารย์ที่ปรึกษาตรวจสอบอีกครั้งหนึ่งเพื่อปรับปรุงแก้ไขก่อนนำไปทดลองใช้

การเก็บรวบรวมข้อมูล

การศึกษาวิจัยในครั้งนี้ผู้วิจัยได้ศึกษาข้อมูลปฐมภูมิและข้อมูลทุติยภูมิ เพื่อช่วยสร้างความเข้าใจพื้นฐานในการวิจัยช่วยออกแบบสอบถาม ตลอดจนช่วยให้ได้ข้อมูลที่สมบูรณ์ยิ่งขึ้นโดยมีแหล่งที่มาของข้อมูล ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1. ข้อมูลทุติยภูมิ (Secondary Data) เป็นการ ค้นหาข้อมูลจากเอกสาร วารสารที่สามารถอ้างอิงได้ ผลงานวิจัยต่าง ๆ ที่เกี่ยวข้องรวมถึงแหล่งข้อมูลทางอินเทอร์เน็ต เพื่อประกอบการสร้างแบบสอบถาม

2. ข้อมูลปฐมภูมิ (Primary Data) เป็นข้อมูลที่ได้จากแบบสอบถามเก็บข้อมูลจากกลุ่มตัวอย่าง จำนวน 105 คน ซึ่งจะดำเนินการเก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างตามที่ได้กำหนดไว้จนครบตามจำนวนโดยขอความร่วมมือจากกลุ่มตัวอย่างที่เป็นคนปั่นจักรยานชาวไทยที่อาศัยอยู่ในกรุงเทพมหานครและอยู่ในเขตลาดกระบัง

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

1. ตรวจสอบข้อมูล (Editing) ผู้วิจัยตรวจสอบความสมบูรณ์ของการตอบแบบสอบถาม โดยแยกแบบสอบถามที่ไม่สมบูรณ์ออก
2. การลงรหัส (Coding) นำแบบสอบถามที่ถูกต้องเรียบร้อยแล้ว มาลงรหัสตามที่ได้กำหนด รหัสไว้ล่วงหน้า
3. การประมวลผลข้อมูลที่ลงรหัสแล้ว ได้นำมาบันทึกโดยใช้เครื่องไมโครคอมพิวเตอร์ เพื่อการประมวลผลข้อมูลซึ่งใช้โปรแกรมสำเร็จรูปเพื่อการวิจัยทางสังคมศาสตร์ (R-Studio)

ในวิจัยนี้มีข้อมูลทั้งหมด 94 ข้อมูล ดังตาราง 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูล

Sex	Age	Weight	Height	dis30min	freq	hr
1	27	88	181	12	3	115.8
1	24	76	173	25	3	117.6
1	35	76	170	10	2	111
1	26	68	180	15	3	170
1	25	67	169	15	1	168

ให้ sex แทนเพศโดยกำหนดให้ 1 คือเพศชาย 2คือเพศหญิง

ให้ age แทนอายุตั้งแต่ 20 ถึง 80 ปี

ให้ weight แทนน้ำหนัก

ให้ height แทนส่วนสูง

ให้ dis30min แทนระยะทางในการปั่นครบ30นาที

ให้ freq แทนความถี่ในการปั่นจักรยานต่อสัปดาห์

ให้ hr แทนอัตราการเต้นของหัวใจขณะเหนื่อย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ขั้นตอนการสร้างแบบจำลอง

วิธีที่นำมาใช้ในงานวิจัยนี้มี 3 วิธีด้วยกันคือวิธีการวิเคราะห์ถดถอยเชิงพหุคูณ ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย และ วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม ซึ่งผู้วิจัยมุ่งเน้นในการศึกษาวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมโดยทำการแบ่งข้อมูลออกเป็น 2 ส่วนคือชุดข้อมูลที่ใช้สำหรับฝึกฝนโดยกำหนดจำนวนข้อมูลเป็น 80% ของข้อมูลทั้งหมด และอีกส่วนหนึ่งคือส่วนของชุดข้อมูลทดสอบจะถูกแบ่งออกเป็น 20% ของข้อมูลทั้งหมด ซึ่งทั้ง 3 วิธีที่กล่าวมานี้จะใช้ชุดข้อมูลเดียวกันทั้งหมด มีแผนงานวิจัยดังรูปที่ 3.6 ก่อนที่จะคำนวณในโปรแกรมควรที่จะนำเข้า Packages ที่จำเป็นเสียก่อนซึ่งประกอบไปด้วย

1. e1071 ใช้สำหรับในกระบวนการ SVR
2. GA ใช้สำหรับกระบวนการ GA
3. Readxl ใช้สำหรับนำเข้าไฟล์จาก Excel
4. ggplot2 ใช้สำหรับแสดงกราฟ

```
library(e1071)
library(ggplot2)
library(readxl)
library(GA)
```

รูปที่ 3.1 แสดง Package ที่นำเข้า

3.2.1 การวิเคราะห์การถดถอยเชิงพหุคูณ

ในการวิเคราะห์ข้อมูลจากแบบสอบถาม ผู้วิจัยได้นำข้อมูลมาหาค่าสหสัมพันธ์เพื่อหาความสัมพันธ์ระหว่างตัวแปรต้นกับตัวแปรตาม

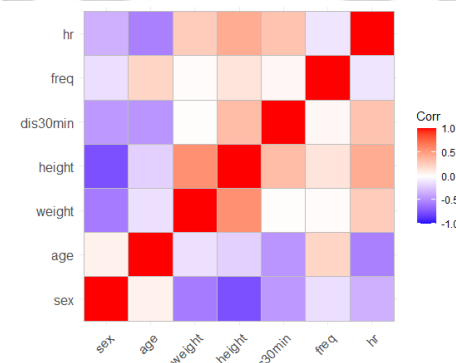
การวิเคราะห์ข้อมูล

1. การวิเคราะห์เพื่อสร้างสมการถดถอยหรือสมการพยากรณ์ปัจจัยที่มีผลต่อเวลาในการปั่นจักรยานโดยหาค่าสหสัมพันธ์ระหว่างตัวแปรตาม (Response , Dependent variable , Y) และตัวแปรต้น (Predictor, Independent variable, X) มากกว่าหนึ่งตัว โดยมี Correlation Coefficient (r) หรือ ค่าสัมประสิทธิ์สหสัมพันธ์เป็นตัวบ่งชี้ถึงความสัมพันธ์ซึ่งค่าสัมประสิทธิ์
- เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สหสัมพันธ์นี้จะมีค่าอยู่ระหว่าง -1.0 ถึง +1.0 ซึ่งหากมีค่าใกล้ -1.0 นั้นหมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันอย่างมากในเชิงตรงกันข้ามหากมีค่าใกล้ +1.0 นั้นหมายความว่า ตัวแปรทั้งสองมีความสัมพันธ์กันโดยตรงอย่างมาก และหากมีค่าเป็น 0 นั้นหมายความว่า ตัวแปรทั้งสองตัวไม่มีความสัมพันธ์ต่อกัน

2. นำตัวแปรที่เลือกนำไปวิเคราะห์เพื่อหาแบบจำลองทางคณิตศาสตร์โดยใช้ การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis)

3. ผู้วิจัยใช้ โปรแกรม R-studio มาช่วยในการวิเคราะห์



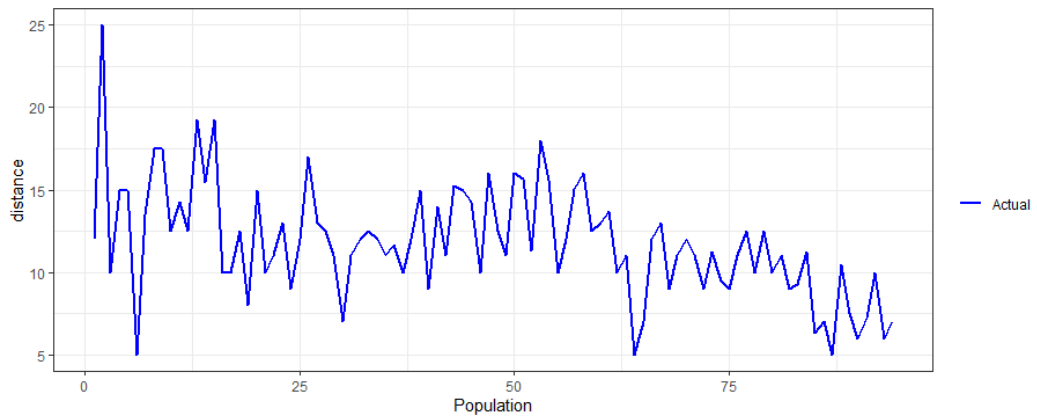
รูปที่ 3.2 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรต่างๆ

3.2.1.1 ทำการนำเข้าข้อมูลจากไฟล์ Excel

	sex	age	Disease	weight	height	type	sysptom	distance(km)	time(hr)
1	1	27	7	88.0	181	2	2	12.0	0.50
2	1	24	0	76.0	173	1	0	50.0	1.00
3	1	35	0	76.0	170	2	0	40.0	2.00
4	1	26	0	68.0	180	2	0	30.0	1.00
5	1	25	0	67.0	169	2	0	30.0	1.00
6	1	25	0	100.0	178	2	0	200.0	2.00
7	1	34	1	113.0	179	2	1	35.0	1.30
8	1	26	0	83.5	175	2	0	70.0	2.00
9	1	30	8	64.0	178	2	1	70.0	2.00
10	1	35	0	80.0	175	2	0	100.0	4.00
11	1	37	0	70.0	174	2	0	50.0	1.75
12	1	37	1	62.0	173	2	1	25.0	1.00
13	1	33	0	75.0	168	2	1	50.0	1.30
14	1	30	0	63.0	164	9	0	23.5	0.76
15	1	35	0	85.0	167	2	0	50.0	1.30

รูปที่ 3.3 แสดงการนำเข้าข้อมูลในโปรแกรม R

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.4 แสดงข้อมูลระยะทางจริง

3.2.2 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย

วิเคราะห์ด้วยโปรแกรม R เนื่องจากได้นำเข้าข้อมูลในโปรแกรมแล้วผู้วิจัยได้นำข้อมูลเดิมมาใช้กับวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression : SVR) ซึ่งสามารถแสดงผลที่ได้ดังรูปที่ 3.5

```
Call:
svm(formula = dis1 ~ sex1 + age1 + we1 + he1 + freq1 + HR1, data = (data1))

Parameters:
  SVM-Type:  eps-regression
  SVM-Kernel: radial
    cost:    1
  gamma:    0.1666667
  epsilon:  0.1

Number of Support Vectors: 80
```

รูปที่ 3.5 แสดงผลลัพธ์จากวิธี SVR

โดยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ในโปรแกรมนี้สามารถค้นหาค่าพารามิเตอร์ที่เหมาะสมได้ สามารถกำหนดเคอร์เนลที่เหมาะสมกับข้อมูลได้ ในที่นี้กำหนดค่าเคอร์เนลคือ Radial Basis Function (RBF) ให้ค่า *Cost* (C) คือ 1 ค่า *Gamma* (γ) คือ 0.1666667 และ ค่า *Epsilon* (ϵ) คือ 0.1 มีจำนวน Support Vector ทั้งหมด 80 ตัว

3.2.3 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิง

พันธุกรรม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากการที่ได้ศึกษาเกี่ยวกับวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย จะได้ว่ามีพารามิเตอร์ที่สำคัญอยู่ 3 ตัว คือ *Cost (C)* *Gamma (γ)* และ *Epsilon (ϵ)* ซึ่งทั้ง 3 ตัวนี้มีผลต่อการพยากรณ์ ผู้วิจัยจึงต้องการที่จะค้นหาค่าโดยนำวิธีเชิงพันธุกรรม มาช่วยในการค้นหาค่าพารามิเตอร์ที่เหมาะสม แล้วจะนำค่าพยากรณ์ที่ได้ไปเปรียบเทียบกับวิธี อื่นๆที่ได้ศึกษามาก่อนหน้านี้

ขั้นตอนวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)

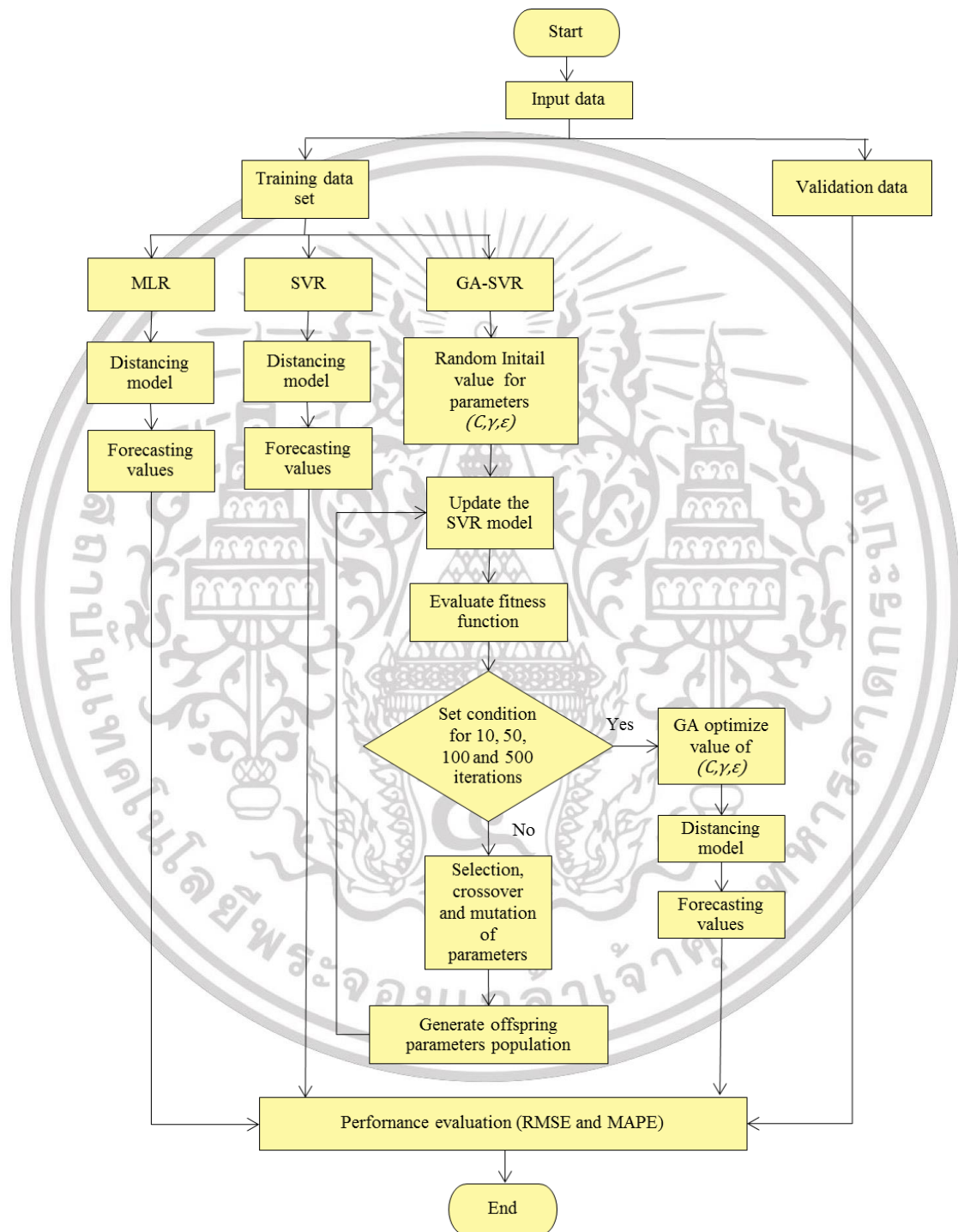
1. เริ่มจากกำหนดค่าเริ่มต้นแบบสุ่มของ C, γ, ϵ โดยค่า C คือค่าคงที่ซึ่งเป็นพารามิเตอร์สำหรับกำหนดปริมาณความผิดพลาด γ คือค่าพารามิเตอร์ของฟังก์ชันเคอร์เนลมีค่าขึ้นอยู่กับความเหมาะสม ϵ แทนระยะเส้นขอบที่ครอบคลุมจุดข้อมูลมากที่สุด โดยกำหนดค่าประชากรคือ 50 และกำหนดค่า C มีค่าอยู่ระหว่าง 0.0001 ถึง 3 ค่า γ มีค่าอยู่ระหว่าง 0.001 ถึง 2 และ ϵ มีค่าอยู่ระหว่าง 0.01 ถึง 2
2. นำค่า C, γ, ϵ ที่ได้ นำเข้าสู่แบบจำลองวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย
3. คำนวณค่า fitness function โดยในงานวิจัยนี้กำหนดค่า fitness function วัดจากค่า RMSE
4. กำหนดการหยุดการทำซ้ำของโปรแกรมโดยในครั้งแรกกำหนดการทำซ้ำ 10 รอบ ถ้ายังทำซ้ำไม่ถึง 10 จะเข้าสู่ขั้นตอนต่อไปแต่หากครบ 10 รอบแล้วจะได้ค่า C, γ, ϵ เหมาะสมที่สุด
5. การสร้างโครโมโซมรุ่นต่อไปผ่านการ คัดเลือก (Selection) สลับสายพันธ์ (Crossover) และ ข้ามสายพันธ์ (Mutation)
6. จะได้ประชากรรุ่นใหม่จากนั้นจึงนำไปอัปเดต ในแบบจำลอง SVR
7. เมื่อทำครบจำนวนรอบและได้ C, γ, ϵ เหมาะสมที่สุดแล้วจากนั้นจะได้เป็นแบบจำลองสำหรับพยากรณ์ระยะทางการบินจักรยานเพื่อนำไปหาค่าพยากรณ์ระยะทาง
8. นำข้อมูลที่ใช้สำหรับทดสอบมาทดสอบแบบจำลอง

การปรับปรุงค่าด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)

ในส่วนของการปรับปรุงค่าในขั้นตอนแรกผู้วิจัยได้ทำการทำซ้ำ 10 รอบ มีการวัดค่าความแม่นยำเปรียบเทียบกับวิธีอื่นๆ หลังจากนั้นจึงทำซ้ำ 50 รอบ 100 รอบ และ 500 รอบและตรวจสอบความแม่นยำของการพยากรณ์อีกครั้ง

3.3 แผนงานวิจัย

รูปที่ 3.6 แสดงแผนงานวิจัยตั้งแต่เริ่มต้นด้วยการนำเข้าข้อมูลแล้วหาแบบจำลองจนสุดท้ายนำแบบจำลองที่ได้มาเปรียบเทียบความแม่นยำในการพยากรณ์



รูปที่ 3.6 แผนงานวิจัย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการวิจัยและอภิปรายผล

ในบทนี้จะแบ่งการแสดงผลออกเป็น 2 ส่วน ส่วนที่หนึ่งคือผลการวิจัยและการอภิปรายผล พยากรณ์ระยะทางที่เหมาะสมในการปั่นจักรยานและผลการวิจัยและการอภิปรายผลหลังจากมีการ ปรับปรุงค่าพารามิเตอร์การพยากรณ์ระยะทางที่เหมาะสมในการปั่นจักรยาน

4.1 ผลการวิจัยและการอภิปรายการพยากรณ์ระยะทางที่เหมาะสมในการปั่นจักรยาน

ในส่วนนี้จะประกอบไปด้วย 3 วิธีในการพยากรณ์ระยะทางการปั่นจักรยานวิธีแรกคือ การ วิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression : MLR) วิธีที่สองคือวิธีซัพพอร์ต เวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Regression : SVR) และวิธีสุดท้ายคือวิธีซัพ พอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (Support vector regression-Based Genetic Algorithm : GA-SVR) จากแบบจำลองสำหรับสองวิธีแรกจะได้ค่า พยากรณ์ของแต่ละวิธีดังรูปที่ 4.1-4.2 แสดงตัวอย่างค่าพยากรณ์ที่ได้

	predic2	rang	sex	age	weight	height	dis30min	freq	hr
1	14.347384	1	1	27	88.0	181.0	12.00000	3	115.8
2	15.952446	2	1	24	76.0	173.0	25.00000	3	117.6
3	14.503003	3	1	35	76.0	170.0	10.00000	2	111.0
4	16.137438	4	1	26	68.0	180.0	15.00000	3	170.0
5	16.076175	5	1	25	67.0	169.0	15.00000	1	168.0
6	13.039619	6	1	25	100.0	178.0	5.00000	7	195.0
7	10.610909	7	1	34	113.0	179.0	13.46154	4	145.0
8	14.743816	8	1	26	83.5	175.0	17.50000	6	175.0
9	16.046291	9	1	30	64.0	178.0	17.50000	5	198.0
10	13.617404	10	1	35	80.0	175.0	12.50000	5	204.0
11	14.346516	11	1	37	70.0	174.0	14.28571	2	175.0
12	16.160679	12	1	37	62.0	173.0	12.50000	5	110.0
13	14.642756	13	1	33	75.0	168.0	19.23077	4	160.0
14	15.895900	14	1	30	63.0	164.0	15.46053	3	190.0
15	13.154027	15	1	35	85.0	167.0	19.23077	2	150.0
16	14.793693	16	1	39	70.0	170.0	10.00000	3	109.0
17	12.490050	17	1	38	83.0	165.0	10.00000	1	191.0
18	12.941081	18	1	39	80.0	165.0	12.50000	3	190.0
19	10.862043	19	2	30	70.0	158.0	8.00000	1	114.0
20	15.275345	20	2	20	43.0	167.0	15.00000	3	120.0

รูปที่ 4.1 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี MLR

	predic1	rang	sex	age	weight	height	dis30min	freq	hr
1	12.344150	1	1	27	88.0	181.0	12.00000	3	115.8
2	14.269358	2	1	24	76.0	173.0	25.00000	3	117.6
3	12.821283	3	1	35	76.0	170.0	10.00000	2	111.0
4	15.233827	4	1	26	68.0	180.0	15.00000	3	170.0
5	14.656181	5	1	25	67.0	169.0	15.00000	1	168.0
6	10.801334	6	1	25	100.0	178.0	5.00000	7	195.0
7	13.115592	7	1	34	113.0	179.0	13.46154	4	145.0
8	13.906957	8	1	26	83.5	175.0	17.50000	6	175.0
9	14.627894	9	1	30	64.0	178.0	17.50000	5	198.0
10	12.843726	10	1	35	80.0	175.0	12.50000	5	204.0
11	14.393235	11	1	37	70.0	174.0	14.28571	2	175.0
12	15.031304	12	1	37	62.0	173.0	12.50000	5	110.0
13	15.424661	13	1	33	75.0	168.0	19.23077	4	160.0
14	15.564520	14	1	30	63.0	164.0	15.46053	3	190.0
15	11.877256	15	1	35	85.0	167.0	19.23077	2	150.0
16	14.379312	16	1	39	70.0	170.0	10.00000	3	109.0
17	10.582971	17	1	38	83.0	165.0	10.00000	1	191.0
18	12.599271	18	1	39	80.0	165.0	12.50000	3	190.0
19	8.345604	19	2	30	70.0	158.0	8.00000	1	114.0
20	14.654943	20	2	20	43.0	167.0	15.00000	3	120.0

รูปที่ 4.2 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี SVR

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.1 ผลลัพธ์ที่ได้จากโปรแกรมวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม

ในกระบวนการครั้งแรกแสดงการทำซ้ำ 10 รอบได้ผลดังรูปที่ 4.3 โดยมีจำนวน Support Vector จำนวน 47

```
Call:
svm(formula = dis30min ~ sex + age + weight +
height + freq + hr, data = (datatr), type = "eps-regression",
kernel = "radial", cost = 1.515713, gamma = 0.1104122,
epsilon = 0.2965661)

Parameters:
SVM-Type: eps-regression
SVM-Kemel: radial
cost: 1.515713
gamma: 0.1104122
epsilon: 0.2965661

Number of Support Vectors: 47
```

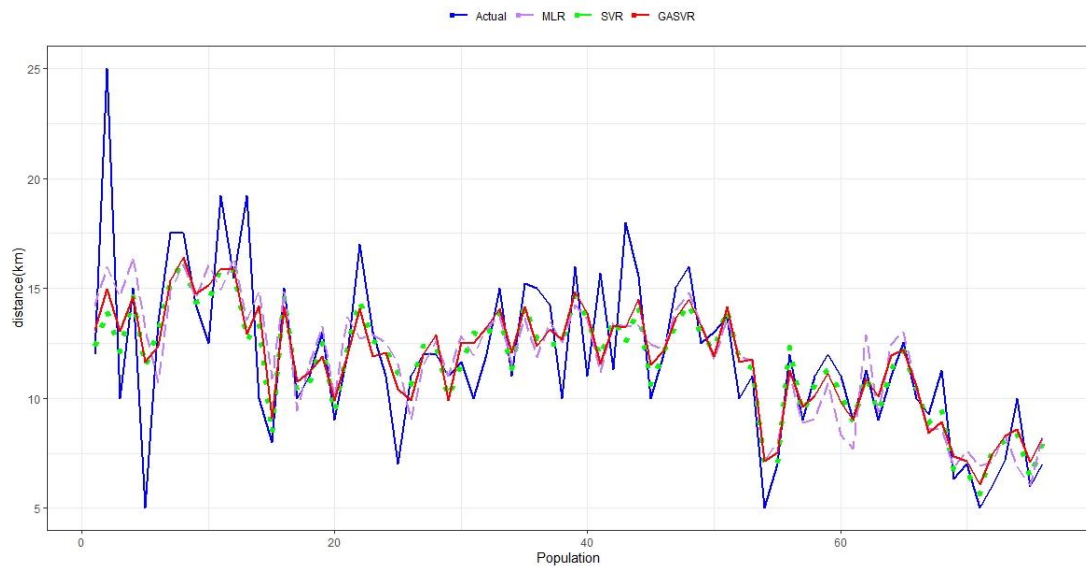
รูปที่ 4.3 แสดงการทำซ้ำ 10 รอบ

	predic3	sex	age	weight	height	dis30min	freq	hr
1	12.144977	1	27	88.0	181.0	12.00000	3	115.8
2	21.636628	1	24	76.0	173.0	25.00000	3	117.6
3	9.996087	1	35	76.0	170.0	10.00000	2	111.0
4	15.150133	1	26	68.0	180.0	15.00000	3	170.0
5	14.915716	1	25	67.0	169.0	15.00000	1	168.0
6	7.101434	1	25	100.0	178.0	5.00000	7	195.0
7	13.311281	1	34	113.0	179.0	13.46154	4	145.0
8	17.349815	1	26	83.5	175.0	17.50000	6	175.0
9	17.222387	1	30	64.0	178.0	17.50000	5	198.0
10	12.649305	1	35	80.0	175.0	12.50000	5	204.0
11	14.428038	1	37	70.0	174.0	14.28571	2	175.0
12	12.757341	1	37	62.0	173.0	12.50000	5	110.0
13	19.380120	1	33	75.0	168.0	19.23077	4	160.0
14	15.563458	1	30	63.0	164.0	15.46053	3	190.0
15	18.928250	1	35	85.0	167.0	19.23077	2	150.0
16	9.980019	1	39	70.0	170.0	10.00000	3	109.0
17	9.977384	1	38	83.0	165.0	10.00000	1	191.0
18	12.558729	1	39	80.0	165.0	12.50000	3	190.0
19	8.151565	2	30	70.0	158.0	8.00000	1	114.0
20	14.848687	2	20	43.0	167.0	15.00000	3	120.0

รูปที่ 4.4 หน้าจอแสดงผลตัวอย่างค่าพยากรณ์ที่ได้ในโปรแกรมด้วยวิธี GA-SVR

เมื่อได้ค่าผลลัพธ์ของทั้ง 3 วิธีแล้วหลังจากนั้นจึงนำมาแสดงด้วยกราฟเปรียบเทียบเพื่อดูแนวโน้มของแต่ละแบบจำลองโดยกำหนดให้เส้นสีน้ำเงินแทนข้อมูลจริง สีม่วงแทนแบบจำลองการวิเคราะห์ถดถอยเชิงพหุคูณ (MLR) สีเขียวแทนแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการเอกซารันเป็นเอกซารันที่สวอนไวสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ถดถอย (SVR) และสี่แดงแทนแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR)



รูปที่ 4.5 เปรียบเทียบทั้ง 3 วิธี

จากตารางที่ 4.1 แสดงเวลาในการคำนวณโดยใช้คอมพิวเตอร์รุ่น OS Intel Core i5-5200U, 4 GB RAM, 64 bit จากผลลัพธ์ที่ได้จะเห็นว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ใช้เวลาน้อยที่สุดและมีการวัดค่าคลาดเคลื่อนเปรียบเทียบกัน จะเห็นได้ว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมให้ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำที่สุดคือ 1.529 วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยให้ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 1.622 และวิธีการวิเคราะห์ถดถอยเชิงพหุคูณให้ค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 1.630 และแสดงเปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยของทั้ง 3 วิธีจะเห็นได้ว่าวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยให้เปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยต่ำที่สุดคือ 0.110 คิดเป็น 11.0% วิธีวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรมให้เปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยคือ 0.113 คิดเป็น 11.3% และวิธีการวิเคราะห์ถดถอยเชิงพหุคูณให้เปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยคือ 0.12 คิดเป็น 12%

ตารางที่ 4.1 แสดงผลเปรียบเทียบ ทั้ง 3 วิธี

Method	Time (minutes)	RMSE	MAPE
MLR	0.010	1.630	0.120
SVR	0.001	1.622	0.110
GA-SVR	0.35	1.529	0.113

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 ผลการวิจัยและการอภิปรายผลหลังจากมีการปรับปรุงค่าพารามิเตอร์

หลังจากที่รู้ผลลัพธ์ของการทำซ้ำ 10 รอบแล้ว ในขั้นต่อไปจะทำการทำซ้ำ 50 100 และ 500 รอบได้ผลดังรูปที่ 4.6-4.8

```
Call:
svm(formula = dis30min ~ sex + age + weight + height + freq +
  hr, data = (datatr), type = "eps-regression", kernel = "radial",
  cost = 1.211957, gamma = 0.09932316, epsilon = 0.2899838)

Parameters:
  SVM-Type:  eps-regression
  SVM-Kernel: radial
    cost:    1.211957
   gamma:   0.09932316
  epsilon:  0.2899838

Number of Support Vectors: 46
```

รูปที่ 4.6 แสดงการทำซ้ำ 50 รอบ

```
Call:
svm(formula = dis30min ~ sex + age + weight + height + freq +
  hr, data = (datatr), type = "eps-regression", kernel = "radial",
  cost = 1.294484, gamma = 0.07818789, epsilon = 0.04721787)

Parameters:
  SVM-Type:  eps-regression
  SVM-Kernel: radial
    cost:    1.294484
   gamma:   0.07818789
  epsilon:  0.04721787

Number of Support Vectors: 70
```

รูปที่ 4.7 แสดงการทำซ้ำ 100 รอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```

Call:
svm(formula = dis30min ~ sex + age + weight + height + freq +
hr, data = (datatr), type = "eps-regression", kernel = "radial",
cost = 1.891269, gamma = 0.04858175, epsilon = 0.02312592)

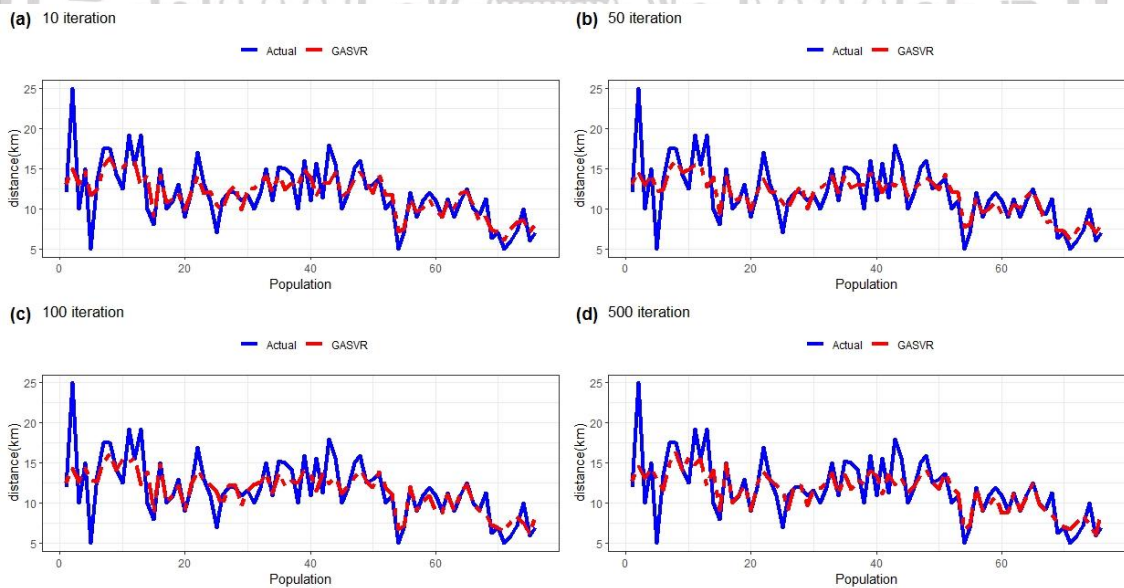
Parameters:
  SVM-Type:  eps-regression
  SVM-Kernel: radial
    cost:    1.891269
   gamma:   0.04858175
  epsilon:  0.02312592

Number of Support Vectors: 72

```

รูปที่ 4.8 แสดงการทำซ้ำ 500 รอบ

หลังจากที่ทำซ้ำ 10 50 100 และ 500 รอบแล้วนำมาแสดงด้วยกราฟจะได้ดังรูปที่ 4.9 โดยให้ (a) แทนการทำซ้ำ 10 รอบ (b) แทนการทำซ้ำ 50 รอบ (c) แทนการทำซ้ำ 100 รอบและ (d) แทนการทำซ้ำ 500 รอบ และให้เส้นสีแดงแทนค่าพยากรณ์วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม (GA-SVR) ให้สีน้ำเงินแทนค่าจริง



รูปที่ 4.9 เปรียบเทียบการทำซ้ำ 10 50 100 และ 500 รอบ

การค้นหาค่าพารามิเตอร์ด้วยวิธีเชิงพันธุกรรม (GA) จะเก็บค่าพารามิเตอร์ทั้ง 3 ค่าที่เหมาะสมที่สุดในการทำซ้ำ 10 50 100 และ 500 รอบ แสดงได้ในตารางที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 แสดงค่าพารามิเตอร์ที่เหมาะสม

Iterations	C	ϵ	γ
10	1.516	0.110	0.297
50	1.211	0.099	0.290
100	1.294	0.078	0.047
500	1.891	0.049	0.023

4.2.1 สรุปผล

จากการรวบรวมข้อมูลจากระยะทางปั่นจักรยานของแต่ละคนในเขตลาดกระบัง กรุงเทพมหานคร โดยรวบรวมข้อมูลด้วยแบบสอบถาม แบบจำลองการพยากรณ์ถูกสร้างขึ้นโดยใช้เทคนิคที่แตกต่างกันสามแบบ ได้แก่แบบจำลองการวิเคราะห์การถดถอยเชิงพหุคูณ วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยและวิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม ทำการเลือกแบบจำลองที่ดีที่สุดโดยการวัดประสิทธิภาพของแบบจำลองโดยการวัดความคลาดเคลื่อนของ RMSE และ MAPE ในส่วน GA-SVR จะใช้ GA เพื่อค้นหาพารามิเตอร์สามตัว (C, ϵ, γ) และนำค่าพารามิเตอร์ไปใส่ในแบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยเพื่อปรับปรุงแบบจำลองและทำซ้ำ 10, 50, 100 และ 500 รอบ และวัดข้อผิดพลาดเพื่อดูประสิทธิภาพของแบบจำลองผลที่ได้คือ การทำซ้ำ 500 ครั้งทำให้เกิดข้อผิดพลาดต่ำสุดคือมีค่ารากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 1.314 และ เปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย 9.1% แต่เนื่องจากการทำซ้ำ 500 รอบใช้เวลาในการคำนวณของโปรแกรมเท่ากับ 10.02 นาที เทียบกับการทำซ้ำ 100 รอบ ซึ่งให้เวลาในการคำนวณของโปรแกรมเพียง 2 นาที แต่มีความคลาดเคลื่อนใกล้เคียงกัน ดังนั้นการวนซ้ำ 100 รอบก็เพียงพอแล้ว จะได้ไม่ใช้เวลาในการคำนวณมากเกินไป

ตารางที่ 4.3 แสดงผลการวัดค่าคลาดเคลื่อนกับเวลาที่ใช้ในการคำนวณ

Method	Time (minutes)	RMSE	MAPE
GA-SVR			
10 iterations	0.35	1.529	0.113
50 iterations	1.03	1.501	0.108
100 iterations	2.00	1.389	0.094
500 iterations	10.02	1.314	0.091

แบบจำลองการพยากรณ์ระยะทางในการปั่นจักรยานที่กล่าวถึงข้างต้นมีจุดมุ่งหมายเพื่อทำนายระยะการปั่นจักรยานของบุคคลที่มีข้อจำกัดทางกายภาพหรือผู้สูงอายุ โดยการเก็บตัวอย่างสำหรับการทดสอบ และสามารถช่วยให้นักปั่นจักรยานกำหนดขีดจำกัดว่าควรปั่นจักรยานครั้งละกี่กิโลเมตรใน 30 นาที แบบจำลองนี้ยังช่วยลดความเสี่ยงของการเจ็บป่วยสำหรับผู้ที่มีข้อจำกัดด้านสุขภาพร่างกาย และช่วยส่งเสริมสุขภาพที่ดีและความเป็นอยู่ที่ดีด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการวิจัย

จากการดำเนินผลการวิจัยเกี่ยวกับการพยากรณ์ระยะทางที่เหมาะสมกับการปั่นจักรยาน โดยใช้วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอยผสมผสานกับขั้นตอนวิธีเชิงพันธุกรรม ได้ข้อสรุปดังต่อไปนี้

5.1 สรุปผลวิจัย

จากการที่ผู้คนสนใจการออกกำลังกายมากขึ้นทำให้ยอดขายอุปกรณ์ออกกำลังกายเพิ่มสูงขึ้น ในงานวิจัยนี้ผู้วิจัยจึงเลือกการออกกำลังกายด้วยจักรยานอยู่กับที่ซึ่งเป็นอุปกรณ์ออกกำลังกายที่คนทุกวัยสามารถใช้ได้ โดยงานวิจัยนี้มุ่งเน้นไปที่บุคคลที่มีข้อจำกัดทางด้านร่างกายหรือผู้สูงอายุเนื่องจากบุคคลกลุ่มนี้ต้องได้รับความใส่ใจเป็นพิเศษ ผู้วิจัยทำการพยากรณ์ระยะทางการปั่นจักรยานที่เหมาะสมสำหรับแต่ละคน โดยในงานวิจัยนี้มีพารามิเตอร์ที่สำคัญได้แก่ เพศ อายุ น้ำหนัก ส่วนสูง ความถี่ในการปั่นจักรยานต่อสัปดาห์และอัตราการเต้นของหัวใจ โดยเก็บรวบรวมข้อมูลจากการทำแบบสอบถามบุคคลที่มีอายุตั้งแต่ 20-80 ปี

จากข้อมูลระยะทางการปั่นจักรยานของแต่ละคนในเขตลาดกระบัง กรุงเทพมหานคร โดยสร้างตัวแบบการพยากรณ์ด้วยเทคนิคที่แตกต่างกัน 3 วิธี ได้แก่วิธีการวิเคราะห์การถดถอยเชิงพหุคูณ (Multiple Linear Regression : MLR) วิธีซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine : SVM) และ Genetic Algorithm-Based Support Vector Regression (GA-SVR) ในส่วนของวิธี GA-SVR ได้นำวิธีขั้นตอนเชิงพันธุกรรม (GA) มาช่วยในการปรับแต่งค่าพารามิเตอร์ ทั้ง 3 ตัวได้แก่ (C, ϵ, γ) ได้ทำการทำซ้ำ 10 รอบแล้วทำการเลือกตัวแบบที่ดีที่สุดจากทั้ง 3 วิธีด้วยการวัดค่าคลาดเคลื่อนด้วยวิธีรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) และเปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ผลลัพธ์คือแบบจำลอง GA-SVR ได้ค่าความแม่นยำสูงที่สุดจากนั้นจึงนำแบบจำลองไปทำซ้ำ 50 100 และ 500 เพื่อปรับปรุงค่าความแม่นยำในการพยากรณ์ ผลลัพธ์ที่ได้คือการทำซ้ำ 500 รอบให้ค่าคลาดเคลื่อนต่ำที่สุดคือรากของค่าคลาดเคลื่อนกำลังสองเฉลี่ยคือ 1.314 และเปอร์เซ็นต์ค่าคลาดเคลื่อนสมบูรณ์เฉลี่ยคือ 9.1% แต่เนื่องจากการทำซ้ำ 500 รอบใช้เวลาคำนวณของโปรแกรม 10.02 นาที เมื่อเปรียบเทียบกับการทำซ้ำ 100 รอบให้เวลาคำนวณของโปรแกรม 2 นาทีแต่มีค่าความคลาดเคลื่อนใกล้เคียงกัน ดังนั้นไม่มีความจำเป็นที่จะต้องการทำซ้ำถึง 500 ครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้วยตัวแบบการพยากรณ์ระยะทางการปั่นจักรยานที่ได้กล่าวมาข้างต้นมีจุดประสงค์เพื่อพยากรณ์ระยะทางการปั่นจักรยานของบุคคลที่มีข้อจำกัดทางด้านร่างกาย หรือ ผู้สูงอายุ โดยนำกลุ่มตัวอย่างมาทดสอบซึ่งจะช่วยให้ผู้ที่ปั่นจักรยานสามารถกำหนดขอบเขตของตัวเองได้ว่าแต่ละครั้งของการปั่นจักรยานควรจะปั่นกี่กิโลในระยะเวลา 30 นาที แบบจำลองนี้ยังช่วยลดความเสี่ยงจากการล้มป่วยของผู้ที่มีข้อจำกัดทางด้านร่างกายและช่วยเสริมสร้างสุขภาพปอดสำหรับผู้ที่มีประวัติปอดอักเสบ แต่ด้วยข้อมูลที่นำมาวิเคราะห์นั้นเป็นการเก็บข้อมูลเบื้องต้นสำหรับทุกคนทั่วไป ผู้วิจัยคิดว่า การพยากรณ์จะแม่นยำกว่านี้ถ้าสามารถเก็บข้อมูลอื่นๆที่มีผลเช่น ความเร็วขณะปั่นจักรยาน ขนาดวงล้อ จักรยาน ค่าออกซิเจนในเลือดระหว่างก่อนออกกำลังกายและหลังออกกำลังกาย เป็นต้น ในส่วนของข้อมูลผู้วิจัยคิดว่า จะต้องทำการวิเคราะห์จุดข้อมูลที่จะนำมาใช้เพิ่มเติมเพื่อผลลัพธ์ที่ดียิ่งขึ้น ดังนั้นการประยุกต์ใช้แบบจำลองต้องศึกษาและปรับปรุงเพิ่มเติม



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

- [1] Buttar, H.S., Li, T. and Ravi, N. "Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity, and smoking cessation." *Experimental and Clinical Cardiology*. 2005. 10(4):229–249.
- [2] Cronkleton, E. "11 Benefits of Cycling, Plus Safety Tips." Available: <https://www.healthline.com/health/fitness-exercise/cycling-benefits>, 2020.
- [3] Mandal, "A. Just 30 minutes of exercise each day can lower heart disease risk. Online." Available: shorturl.at/cqvLQ.
- [4] Garber, C.E., Blissmer, B., Deschenes, M.R., Franklin, B., Lamonte, M.J., Lee, I.M., Nieman, D.C. and Swain, D.P. "Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults." *Medicine & Science in Sports & Exercise*. 2011. 43(7): 1334-1359.
- [5] พรพนา โยธารักษ์, พรพภา มีเคลือบ และ ศิริรัตน์ คงสังข์. "เกมคณิตศาสตร์กับการพัฒนาสุขภาพ" วิทยาศาสตร์บัณฑิต (คณิตศาสตร์ประยุกต์). สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2561.
- [6] Research and Markets. *Fitness Equipment: COVID-19*. Available: <https://www.researchandmarkets.com>
- [7] Elias, K. Tomaras and Brian, R. MacIntosh. "Less is more: standard warm-up causes fatigue and less warm-up permits greater cycling power output." 2011. 111: 228–235.
- [8] Whiten S., 2014. Team SKY Warm-up. [Online]. Available : <http://britishcyclesport.com/2014/training/team-sky-warm-up/>
- [9] Vapnik, V. "The Nature of Statistical Learning Theory." New York: Springer. 1995.
- [10] Sanguansat, P. "Artificial Intelligence with Machine Learning." Nonthaburi: IDC Premier. 2019.
- [11] Awad, M. and Khanna, R. "Support Vector Machines for Classification." 2015
- [12] Granata, F., Gargano, R. and Marinis, G.D. "Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model." *Water*, 2016, 8:3, 69.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [13] Parrella, F. "Support Vector Machines for Regression." Online. Available: <http://onlinesvr.altervista.org/>.
- [14] รณชัย ชื่นธวัช, กิตติศักดิ์ เกิดประสพ และนิตยาเกิดประสพ. "การพยากรณ์ความต้องการใช้ งานหน่วยจา หน่วยไฟฟ้าด้วย ซัพพอร์ต เวกเตอร์เรกเรชัน แบบตรวจสอบสลบ 3 ส่วน." 2560. 19(1) : 215-232
- [15] Zander, A., Passmore, E., Mason, C. and Rissel, C. Joy, Exercise, Enjoyment, "Getting out: A Qualitative Study of Older People's Experience of Cycling in Sydney, Australia." *Journal of Environmental and Public Health*. 2013. 2013(5) : 547453.
- [16] Titze, S., Geus, B.D., Krenn, P., Bauman, A., Oja, P., Nash, B.R. and Kohlberger, T. "Health benefits of cycling: A systematic review." *Scandinavian Journal of Medicine and Science in Sports*. 2011. 21(4) : 496-509.
- [17] Chuentawat, R., Kerdprasop, K. and Kerdprasop, N. The Demand Forecasting of Electrical Power Distribution Units by Support Vector Regression with 3 Folds Cross-Validation. *Journal of science and technology, Ubon Ratchathani University*. 2017. 19(1).
- [18] Ogcı, G., Demirel, O.F., and Zaim, S. "Forecasting Electricity Consumption with Neural Networks and Support Vector Regression." *Procedia - Social and Behavioral Sciences*. 2012. 58 : 1576-1585.
- [19] Xu, L., Zhang, W., Zhao, J., Cao, Z., Xie, R., Liu, X. and Hu, J. "Support-vector-regression-based prediction of water holdup in horizontal oil-water flow by using a bicircular conductance probe array." *Flow Measurement and Instrumentation*. 2017. 57 : 64-72.
- [20] Yangming, G., Lu, Z., Xiaobin, C., Congbao, R., Zhengjun, Z. and Jiezhong, M. "Time Series Adaptive Online Prediction Method Combined with Modified LS-SVR and AG." *Mathematical Problems in Engineering*. 2012.
- [21] Suzukia, Y., Ibayashia, H., Kanedab, Y. and Minenoa, H. "Proposal to sliding window-based support vector regression-KES2014." *18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*. 2014. 35: 1615-1624.

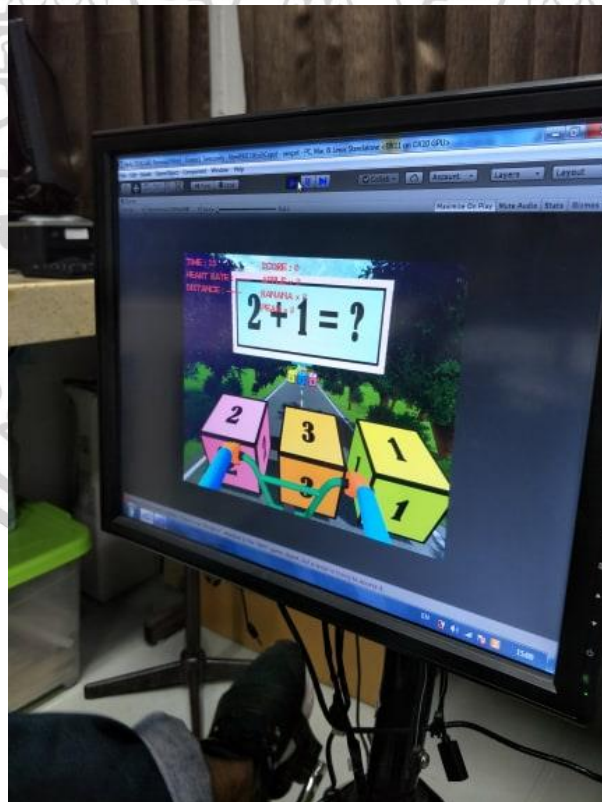
- [22] Yang, C.C. and Shieh, M.D. “A support vector regression based prediction model of affective responses for product form design.” *Computers & Industrial Engineering*. 2010, 59, 682–689
- [23] Zangoeei, M.H., Habibi, J. and Alizadehsani, R. “Disease Diagnosis with a hybrid method SVR using NSGA- II.” *Neurocomputing*. 2014. 136(14–29).
- [24] Tang, X., Wang, L., Chang, J., Chen, J. and Sheng, V. S. “Forecasting model based on information-granulated GA-SVR and ARIMA for producer price index.” 2019.
- [25] Guo, M. L., Li, D., Du, C., Jia, Z., Qin, X., Chen, L., Sheng, L. and Li, H. “Prediction of the Busy Traffic in Holidays Based on GA-SVR.” *Advances in Computer Science and Information Engineering*. 2012. 577-582.
- [26] Kao, Y. S., Nawata, K. and Huang, C. Y. “Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition.” *Mathematics* 2020. 2020. 8(10).
- [27] Liu, S., Tai, H., Ding, Q., Li, D., Xu, L. and Wei, Y. “A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction.” *Mathematical and Computer Modelling*. 2013. 58(3–4): 458-46.
- [28] Ahn, H., Kim, S. and Kim, J. K. “GA-optimized Support Vector Regression for an Improved Emotional State Estimation Model.” *KSI Transactions on Internet and Information Systems*. 2014, 8(6) : 2056-2069
- [29] Lu, C. J., Lee, T. S. and Chiu, C. C. “Financial time series forecasting using independent component analysis and support vector regression.” *Decision Support System*. 2009, 47(2) : 115-125
- [30] Hong, T. H. and Kim, E. M. “The prediction of Purchase Amount of Customers Using Support Vector Regression with Separated Learning Method.” *Journal of Intelligence and Information System*. 16(4) : 213-225



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เครื่องปั่นจักรยานอยู่กับที่ที่ใช้ทดสอบ



เครื่องปั่นจักรยานอยู่กับที่ที่ใช้ทดสอบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางข้อมูลของเพศชาย ช่วงอายุ 20-40 ปี จำนวน 18 คน

	age	weight	height	dis30min	freq	hr
	27	88	181	12.0	3	116
	24	76	173	25.0	3	118
	35	76	170	10.0	2	111
	26	68	180	15.0	3	170
	25	67	169	15.0	1	168
	25	100	178	50.0	7	195
	34	113	179	13.5	4	145
	26	84	175	17.5	6	175
	30	64	178	17.5	5	198
	35	80	175	12.5	5	204
	37	70	174	14.3	2	175
	37	62	173	12.5	5	110
	33	75	168	19.2	4	160
	30	63	164	15.5	3	190
	35	85	167	19.2	2	150
	39	70	170	10.0	3	109
	38	83	165	10.0	1	191
	39	80	165	12.5	3	190
ค่าสูงสุด	39	113	181	50.0	7	204
ค่าต่ำสุด	24	62	164	10.0	1	109
ค่าเฉลี่ย	32	78	172	17	3	160
ส่วนเบี่ยงเบนมาตรฐาน	5	13	5	9	2	34

ตารางข้อมูลของเพศหญิงช่วงอายุ 20-40 ปี จำนวน 14 คน

	age	weight	height	dis30min	freq	hr
	30	70	158	8.0	1	114
	20	43	167	15.0	3	120
	37	70	172	10.0	2	165
	30	63	163	11.0	2	114
	33	46	150	13.0	2	112
	35	74	154	9.0	2	111
	29	46	152	12.0	2	115
	28	52	170	17.0	3	145
	28	50	160	13.0	3	190
	30	52	153	12.5	2	114
	32	52	160	11.0	2	112
	37	52	157	7.0	2	185
	39	76	158	11.0	3	160
	39	55	163	12.0	3	100
ค่าสูงสุด	39	76	172	17	3	190
ค่าต่ำสุด	20	43	150	7	1	100
ค่าเฉลี่ย	32	57	160	12	2	133
ส่วนเบี่ยงเบนมาตรฐาน	5	11	7	3	1	30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางข้อมูลของเพศชาย ช่วงอายุ 41-60 ปี จำนวน 31 คน

	age	weight	height	dis30min	freq	hr
	44	75	175	12.5	1	165
	50	74	171	12.0	2	165
	50	92	173	11.0	1	102
	53	74	163	11.7	2	100
	42	90	175	10.0	3	168
	44	76	172	12.0	2	165
	47	68	171	15.0	3	155
	54	96	174	9.0	1	150
	49	72	180	14.0	2	160
	55	85	170	11.0	5	99
	47	68	168	15.2	3	188
	44	89	172	15.0	3	168
	47	75	170	14.3	4	104
	53	72	173	10.0	3	162
	49	61	168	16.0	2	157
	43	70	175	12.5	4	199
	49	68	169	11.0	3	152
	55	65	166	16.0	5	99
	56	80	175	15.7	3	180
	45	74	168	11.3	5	205
	55	70	170	18.0	7	99
	56	60	168	15.6	1	170
	49	80	170	10.0	2	115
	50	78	177	12.0	3	160
	52	62	170	15.1	2	101
	49	60	165	16.0	3	103
	43	76	170	12.5	2	160
	50	81	180	13.0	3	170
	49	65	173	13.7	3	185
	57	74	174	10.0	4	150
	58	70	186	11.0	2	170
ค่าสูงสุด	58	96	186	18	7	205
ค่าต่ำสุด	42	60	163	9	1	99
ค่าเฉลี่ย	50	74	172	13	3	149
ส่วนเบี่ยงเบนมาตรฐาน	5	9	5	2	1	33

ตารางข้อมูลของเพศหญิงช่วงอายุ 41-60 ปี จำนวน 9 คน

	age	weight	height	dis30min	freq	hr
	42	88	165	5.0	1	149
	41	81	165	7.0	1	145
	41	56	157	12.0	5	170
	45	54	168	13.0	4	160
	50	65	160	9.0	1	102
	56	56	157.5	11.0	1	120
	43	58	160	12.0	2	90
	57	65	156	11.0	3	98
	58	66	173	9.0	3	100
ค่าสูงสุด	58	88	173	13	5	170
ค่าต่ำสุด	41	54	156	5	1	90
ค่าเฉลี่ย	48	65	162	10	2	126
ส่วนเบี่ยงเบนมาตรฐาน	7	12	6	3	2	30

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางข้อมูลของเพศชายช่วงอายุ 61-80 ปี จำนวน 10 คน

	age	weight	height	dis30min	freq	hr
	75	48	170	11.3	4	87
	78	59	162	9.5	3	92
	63	95	165	9.0	2	93
	66	65	169	11.0	7	95
	57	67	174	12.5	5	100
	70	75	175	10.0	2	90
	61	76	178	12.5	7	110
	69	69	173	10.0	6	90
	62	87	171	11.0	7	92
	70	89	167	9.0	2	89
ค่าสูงสุด	78	95	178	13	7	110
ค่าต่ำสุด	57	48	162	9	2	87
ค่าเฉลี่ย	67	73	170	11	5	94
ส่วนเบี่ยงเบนมาตรฐาน	7	14	5	1	2	7

ตารางข้อมูลของเพศหญิงช่วงอายุ 61-80 ปี จำนวน 12 คน

	age	weight	height	dis30min	freq	hr
	67	50	157	9.3	3	92
	65	55	151	11.3	4	92
	74	58	161	6.3	3	87
	71	58	156	7.0	6	89
	80	52	157	5.0	4	85
	69	60	160	10.5	2	92
	70	54	158	7.5	3	90
	79	52	152	6.0	3	86
	62	60	158	7.2	5	102
	68	67	154	10.0	4	93
	77	66	156	6.0	6	88
	63	58	161	7.0	4	95
ค่าสูงสุด	80	67	161	11	6	102
ค่าต่ำสุด	62	50	151	5	2	85
ค่าเฉลี่ย	70	58	157	8	4	91
ส่วนเบี่ยงเบนมาตรฐาน	6	5	3	2	1	5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

MATEMATIKA: MJIAM, 2021, Volume 37, Number 3, 133–144
 © Penerbit UTM Press. All rights reserved

Forecasting A Cycling Distance Based on Personal Health Data with Hybrid GA-SVR Approach

¹Pawnwipa Meeklueb, ²Busayamas Pimpunchat* and ³Benjawan Rodjanadid

^{1,2}Center of Business Analytics and Synthetic Intelligence Strategy, Department of Mathematics
 Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

³School of Mathematics, Institute of Science
 Suranaree University of Technology, Nakhon Ratchasima, 30000, Thailand

*Corresponding author: busayamas.pi@kmitl.ac.th

Article history

Received: 21 August 2021

Received in revised form: 11 November 2021

Accepted: 21 November 2021

Published online: 15 December 2021

Abstract The aim of this study is to forecast the optimal cycling distance of 30 minutes approximately for physically constrained individuals, especially those having a congenital disease or elderly person. Data illustrations are collected from having 94 people who cycled in Ladkrabang by questionnaires. The research focuses on forecasting optimal distances by exploiting Support Vector Regression based on Genetic Algorithm (GA-SVR) that will be employed to improve the accuracy of the numerical study. We offer performance comparisons of our model against Multiple Linear Regression (MLR) and Support Vector Regression (SVR) algorithms. The experimental results demonstrate that GA-SVR outperforms MLR and SVR based on the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). This GA-SVR model is proven to be an effective approach to predict cycling distances.

Keywords Optimal cycling distance; Health; Support Vector Regression; Genetic Algorithm; Multiple Linear Regression.

Mathematics Subject Classification 6211; 62P05; 62J99.

1 Introduction

Human beings are becoming increasingly interested in exercising nowadays since there are various forms of exercise, either with or without equipment. The merit of exercising is keeping the body healthy, but overexertion can lead to injuries. Settling with knowing exercise limits is recommended, and also it can contribute to reducing risks of adverse effects later on. Cycling is a popular exercise for people of all ages, yet some groups have limitations with physical activities i.e. somebody with high blood pressure, an elderly person with diabetes, or groups at risk of developing coronary artery disease. Adequate designs of exercises combined with a healthy diet can reduce the risk of developing the disease [1, 2]. Physical activity is prompted

37:3 (2021) 133–144 | www.matematika.utm.my | eISSN 0127-9602 |

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

for adults aged 18-64 years by performing moderate-intensity aerobic exercise for at least 150 minutes per week, or sets of 30 minutes, five times per week [3]. Much of those with physical limitations can only be flexible 2-3 times a week [4]. According to research and markets [5], the world's leading source for international market research reports and market data, it is stated that fitness equipment sales have gone up to 170%. Since gyms are required to close, it prompts many people to buy exercise equipment, i.e. exercise bikes, treadmills, and rowing machines. In a stationary bicycle, there is usually a screen displaying stats, i.e. the person's heart rate, distance while cycling, etc. It will be better if we can forecast the appropriate distance for each person. This is consistent with the current situation due to the widespread disease of the coronavirus outbreak, making people unable to exercise in a public space such as exercising at home is an alternative.

For earlier studies, exercise can be beneficial effect on the elderly person [6] and cycling reduces the risk of various diseases in both children and adults [7]. The most common forecasting techniques are the Multiple Linear Regression (MLR) and the Support Vector Regression (SVR) methods, which are regression analyses between vectored inputs and output variables using machine learning principles [8]. These methods provide high accuracy [9-12], but the SVR has limitations in its functionality since the user has to configure the parameters accordingly to get an ideally powerful SVR. Therefore, people prefer to use the SVR with other methods to get more efficiency, Yang *et al.* [13], have developed a model that predicts consumers' affective responses (CARs) for producing a form design. The predictive performance of the SVR with RCGA (SVR-RCGA) is compared to that of SVR with 5-fold cross-validation (SVR-5FCV) and a back-propagation neural network (BPNN) with 5-fold cross-validation (BPNN-5FCV). The experimental results using the data sets on mobile phones and electronic scooters show that SVR performs better than BPNN. MLR methods have been proposed by collecting data from cyclists on a cycling track as a sample for forecasting the distance of a stationary bike.

To find the appropriate parameters in the SVR, we propose the Genetic Algorithm based on Support Vector Regression (GA-SVR). GA-SVR can be applied to solve forecasting a wide variety of problems and even enhance it, i.e. Zangoeei *et al.* [14], displayed the diagnosis of disease by the prognosis of four diseases, namely breast cancer, liver disease, hepatitis, and diabetes by utilizing SVR and NSGA-II. The results of the accuracy compared to other methods have found that the application of the SVR method with Non-dominated Sorting Genetic Algorithm II (NSGA-II) has higher accuracy than any other method. Tang *et al.* [15], proposed an unprecedented hybrid model based on a fuzzy information granulation that is integrating models of the GA-SVR and Autoregressive Integrated Moving Average (ARIMA) in the predicting of the Producer Price Index (PPI). Their research showed that the PPI value predicted by this hybrid model is more accurate than that predicted by other models. Guo *et al.* [16], forecasted the loud traffic of Xinjiang in holidays. All of the previous studies can be referred to [17, 18] and other related references. Numerical comparisons of this paper between MLR, SVR, and GA-SVR with measurements of RMSE and the MAPE are discussed in section 3 of this paper. In this paper, we are keen on the optimal cycling distance in 30 minutes so that users might utilize it as a guide to avoid excessive exercise effects that are possibly causing more peril than nifty.

2 Materials and Methods

The variables mentioned above are used for modeling both SVR and GA-SVR methods. There are 94 data in this study, divided into two parts, training data, and validation data. The numbers of training and validation data are divided into 80:20 ratio respectively. This split data is used in MLR, SVR, and GA-SVR methods to compare a model performance.

2.1 Previous Study

In 2019, Yotharak *at el.* [19], had been created the model using the MLR method with the following variables: Cycling distance in 30 minutes, gender, age, weight, height, cycling frequency, and heart rate. All of these variables are found having correlations since of the MLR method. The data can be divided as follows: male aged 20-40 years, female aged 20-40 years, male aged 41-60 years, female aged 41-60 years, male aged 61-80 years and female aged 61-80 years and are then forecasted to get a total of 6 equations. This is because when using the data before sharing to forecast, the accuracy is increasingly low, but after the data is divided, it can be forecasted with better accuracy but is still not decent. Our past study used MLR attempts to model the relationship between the more explanatory variables and a response variable by fitting a linear equation to the observed data. Every value of the independent variable x is associated with a value of the dependent variable y . The multiple linear regression model be shown below

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + b \quad (1)$$

with a_0 as intercept, x_1, x_2, \dots, x_5 instead of ages, weight, height, cycling frequency, and heart rate, successively. y be a distance for 30 minutes, a_1, a_2, \dots, a_5 represent coefficients of the model, and b is a forecast error.

2.2 Data

This study examines the analysis of the optimal distance for cycling. The instrument used for this data collection is questionnaires, which is a collection of data from people who cycled at the "Happy and Healthy Bike Lane" in Ladkrabang, Bangkok, Thailand, with a total of 94 samples, aged between 20-80 years. Table 1, represents examples of the data collected including Gender (let the number 1 reflects the man, 2 with the woman), Age in year, Weight in kilogram, Height in centimeter, Dis30min is the distance for 30 minutes (kilometers), Freq is cycling frequency (times per week) and HR is heart rate (beats per minute) which is recorded at the moment of most exhaustion.

2.3 Support Vector Regression

This part will recall tools of the Support Vector Machine (SVM) that is useful for our main discussions. For developing a Support Vector Regression (SVR) model, the Support Vector Machine (SVM) should be evaluated. In SVM, the largest margin between one class and another one is used as the basis for deciding the optimal hyperplane, and the data vectors at the edge of each class are called support vectors [20], hyperplanes are split into 2 types. Figure 1(a), shows a hyperplane capable of dividing two classes of data correctly, the Hard-Margin

Table 1: Sample information of people riding bicycles

Gender	Age	Weight	Height	Dis30min	Freq	HR
1	27	88	181	12	3	115.8
1	24	76	173	25	3	117.6
1	35	76	170	10	2	111
1	26	68	180	15	3	170
1	25	67	169	15	1	168

method is used. Figure 1(b), shows a hyperplane that cannot properly divide the two classes of data and use of the Soft-Margin method [21].

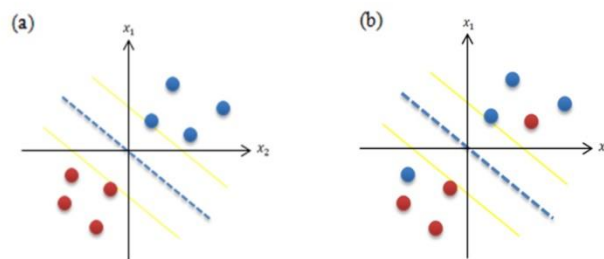


Figure 1: (a) hard-margin SVM and (b) soft-margin SVM

SVM is proposed by Vapnik [22], who thought that theoretically, a linear function f exists to define the nonlinear relationship between the input and output data in a high-dimensional-feature space. Such a method can be used to solve the function concerning ?tting problems [23]. The SVR uses the same principle as SVM for classification, but regression problems. The main idea is always the same: to minimize error, individualizing the hyperplane which maximizes the margin, keeping in mind that part of the error is tolerated [24]. In the case of finding a linear relationship between the input vectors of n ($X \in R^n$) dimensions and the output variables ($y \in R$) of the training data set [8], the SVR regression equation is similar to SVM hyperplane equations because the SVR equation is adapted from SVM. The SVR model is displayed as following

$$f(x) = w^T x + b, \quad (2)$$

where w and b denoted as a weight vector and a bias, respectively. This leads to an optimization problem that minimizes the objective function as

$$\text{minimize } \frac{1}{2} \|w^2\| \quad \text{subject to } y_i - (w^T x_i) - b \leq \varepsilon, \quad (w^T x_i) + b - y_i \leq \varepsilon. \quad (3)$$

It is implicitly assumed that there is a function f that approximates all the pairs (x_i, y_i) with ε precision. In the case where this is not possible, or a specified error can be tolerated, slack variables ξ_i, ξ_i^* can be stated according to the formulation [25] as following

$$\text{minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \text{ subject to } y_i - (w^T x_i) - b \leq \varepsilon + \xi_i, (w^T x_i) + b - y_i \leq \varepsilon + \xi_i^*, \quad (4)$$

where C is a constant which is a parameter for determining the error quantity. In some cases, we cannot directly find the value w . The equation (4) above can solve the dual formulation by applying the Lagrange Multipliers with the Karush-Kuhn-Tucker (KKT) condition. After solving the equation for the value of w , then nonlinear hyper-plane regression functions can be formulated in equation (5) as follows

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b, \quad (5)$$

where α_i and α_i^* are Lagrange's multipliers and $K(x_i, x)$ is a kernel function [25], employed in cases of nonlinear regressions, can be exploited and relegated input vectors to higher dimensions feature space. The kernel function commonly used in SVR formulation as

$$\text{Linear: } K(x_i, x_j) = x_i^T x_j, \quad (6)$$

$$\text{Polynomial: } K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \quad (7)$$

$$\text{Radial Basis Function (RBF): } K(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2}, \gamma > 0, \quad (8)$$

$$\text{Sigmoid: } K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (9)$$

For kernel functions, the parameters γ , r and d in equations (7) until (9), are constants depending on their suitability. For forecasting process, the input data i.e., a distance for 30 minutes and related factors, for instance, gender, age, weight, height, cycling frequency, and heart rate, severally. The forecasting part, SVR model utilized the Radial Basis Function (RBF) kernel function from the previous research [26], which suggests the RBF is usually superior to other kernel functions. The calculation procedure will call the function in R programming. The forecasting result has gotten a cycling distance in 30 minutes. This research evaluates the forecast error values with RMSE and MAPE. We offered a GA-SVR method, (C, ε, γ) as crucial parameters, improve values to achieve a better and highest accuracy than SVR method.

2.4 Genetic Algorithm-based Support Vector Regression

To optimize model based on GA-SVR, a GA is needed to optimize the parameters (C, ε, γ) in the SVR model where C as a penalty coefficient, γ for a kernel function coefficient, and ε is the insensitive coefficient, which has the following main steps [27]:

1. Constructing the chromosome assemblage (C, ε, γ) and formulating the fitness calculation function of the genetic algorithm;
2. Determining the parameters of selection, crossover, mutation, and so on in the GA, and setting the iterative termination condition of the algorithm;
3. Initializing the GA and generating the initialization population;

4. Calculating individual fitness in chromosome populations;
5. Generating next-generation chromosomes through selection, crossover, variation, etc.

Exercise for 30 minutes is a suitable time for exercise, the focus is on using time as one of the limitations in cycling distance forecasting as a guide for those interested in knowing the amount of cycling distance for 30 minutes in kilometers (km). In addition, for those with underlying medical conditions, this model can be used to reduce the risk of injury from excessive exercise. The modeling process for cycling distance of forecasting to compare with MLR, traditional SVR, and GA-SVR models as shown in Figure 2, which is summarized as follows.

1. (Input Data). Call data sets consisting of gender, age, weight, height, cycling frequency per week, heart rate, and distance;
2. Separating (1) into trained data (80%) and validated data (20%);
3. (MLR and SVR). Apply for the same data sets to build the cycling model in optimal distance with MLR and traditional SVR methods. Finally, we can find the forecast data of two methods;
4. (GA-SVR). Three parameters are (C, ε, γ) , for random initial values case, then apply for it for updating the SVR model;
5. (GA-SVR model). As for the GA-SVR model, random initial values of the three parameters (C, ε, γ) then apply them to update the SVR model calculating the fitness function by employing RMSE:
 - a. The stop condition for the calculation is the maximum number of rounds (10, 50, 100, and 500 rounds of repeated sequences). If they do not reach the maximum rounds yet, continue with the calculation inside the loop and through the selection, crossover, and mutation processes;
 - b. The results are parameter sets of a new population. These parameters can be put through an update again to see if they reach the maximum rounds in the set condition;
 - c. If so, bring them out of the loop to get the most optimal parameters. Then, put the obtained parameters by utilizing a GA into the SVR model. The obtained result will be the GA-SVR cycling distance model;
 - d. To find data forecasts of the GA-SVR method.
6. (Measure Performance). To compare the effectiveness of the three models with RMSE and MAPE.

3 Results and Discussions

This part is the main section of creating an effective machine learning model that uses the OS Intel Core i5-5200U, 4 GB RAM, 64 bit for programming of R numerical simulations. To evaluate the performance of the forecasting models, RMSE and MAPE were occupied to evaluate testing performance, which is defined by

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2}. \quad (10)$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right|. \quad (11)$$

From equations (10) and (11), y_t and \hat{y}_t named actual and forecasting values, while n indicated the length of the validation subset. This research compared the performance of the 3 models with RMSE and MAPE error measurements from the validation data, MLR, and SVR as shown in Table 2.

Table 2: Comparisons of the efficiency of MLR and SVR

Method	Time (minutes)	RMSE	MAPE
MLR	0.010	1.630	0.120
SVR	0.001	1.622	0.110

We found that the SVR method had given more accuracy than MLR. The results are decent for acceptance, but the group study was compulsory to be examined with limited health conditions, so the efficiency needs to be improved. As a result, the GA method was used to combine with the SVR to achieve the closest forecast, the comparison results are shown in Figure 3.

The figure shows the performance comparison of the 3 methods with the GA-SVR method is close to original values. In MLR, the correlation of variables was examined before the forecasting process, then the variables relating to each other were taken in forecasting both traditional SVR and GA-SVR shown in the next page. Based on the training data showing the comparison of forecasting performance of 3 models using validation data. It is obtained that the GA-SVR method achieved the best results. Therefore, this method was utilized to test the number of cycles by repeating for 10, 50, 100 and 500 iterations, successively. A radial basis function was applied for a kernel function, and for optimal parameters in the GA method, RMSE is occupied by calculating the fitness function and determining populations with 50 sizes.

In Table 3, the three parameters (C, ε, γ) are values of the GA-SVR process iteration is utilized to forecast the optimal cycling distance. In Figure 4, (a), (b), (c) and (d) are shown for each of 10, 50, 100, and 500 iterations, respectively. Further, GA-SVR algorithms have been implemented into a computer simulation code written in R programming.

The three parameters (C, ε, γ) were obtained at each iteration. Afterwards, these parameters were added into the SVR model. The operating time frame for the GA-SVR method calculation is shown in Table 4, tests with validation data, in seconds and classified into some of 10, 50, 100, and 500 iterations, with the result of 0.35, 1.03, 2.00 and 10.02 minutes respectively.

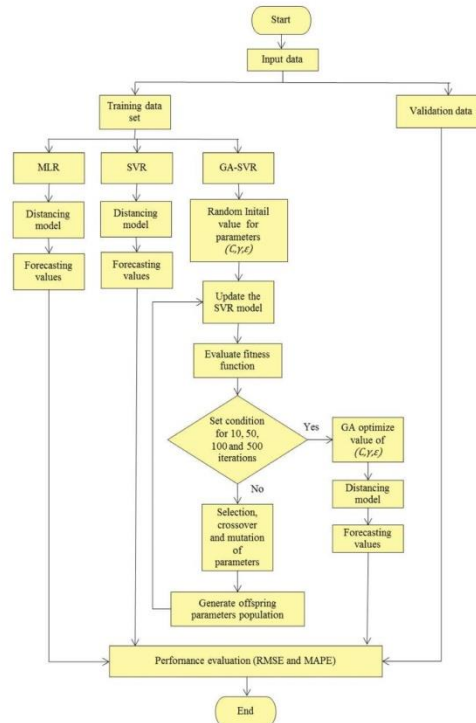


Figure 2: The modeling process for forecasting cycling distance

Remarks are gained that as the number of computations iterated increases, the computation time increases with the GA-SVR method. For comparing between 500 iterations and 100 iterations, we found that 500 iterations have given a lower error value with 10.02 minutes, while 100 iterations have taken time only 2 minutes. Consequently, it can be concluded that 100 iterations are the most effective. The RMSE and MAPE values of the model are calculated as seen on Table 4.

4 Conclusions

The data is gathered from the cycling distance of each person in Ladkrabang, Bangkok by collecting information with the questionnaire. The forecast model is created using three different techniques namely the MLR, SVR, and GA-SVR models. The best and most effective model is selected by measuring the tolerances of RMSE and MAPE. In the GA-SVR section, GA is

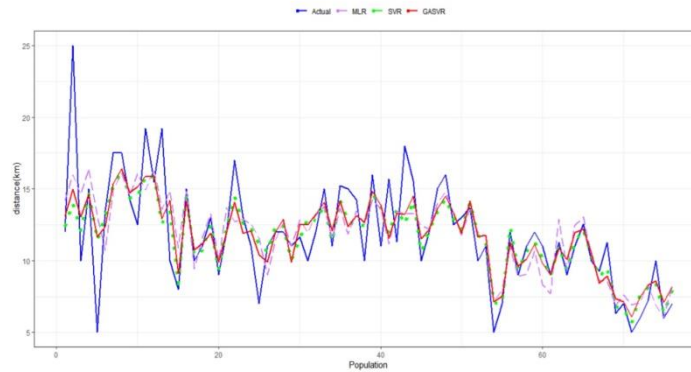


Figure 3: Comparisons between actual and forecast

Table 3: Iteration comparisons of GA-SVR parameters

Iterations	C	ε	γ
10	1.516	0.110	0.297
50	1.211	0.099	0.290
100	1.294	0.078	0.047
500	1.891	0.049	0.023

needed to find the optimal three parameters (C, ε, γ), insert them into the SVR to improve and replicate them for 10, 50, 100 and 500 iterations and measure the error to evaluate the best value. The results 500 iterations yielded the lowest error, RMSE is 1.314 and MAPE is 0.091, but 500 iterations took the program's calculation time of 10.02 minutes, compared to 100 repetitions which gave the program computation time for only 2 minutes, but with similar tolerances. Therefore, the iteration of 100 times is sufficient as well as it will not fill in a variety of times to calculate. The forecasting cycling distance model mentioned above is intended to forecast the cycling distance of a person with physical limitations or the elderly by taking a sample for testing and can help cyclists set limits on how many kilometers each ride should be cycled in 30 minutes. The model also reduces the risks of falling ill for people with physical health limitations and helps to promote good health and well-being as well. The reader is hoping to develop calculating the cycling distance of 30 minutes by adding to the display screen of the stationary bike in the gym for further research. The results would be even better if we could forecast the suitable distance for each person.

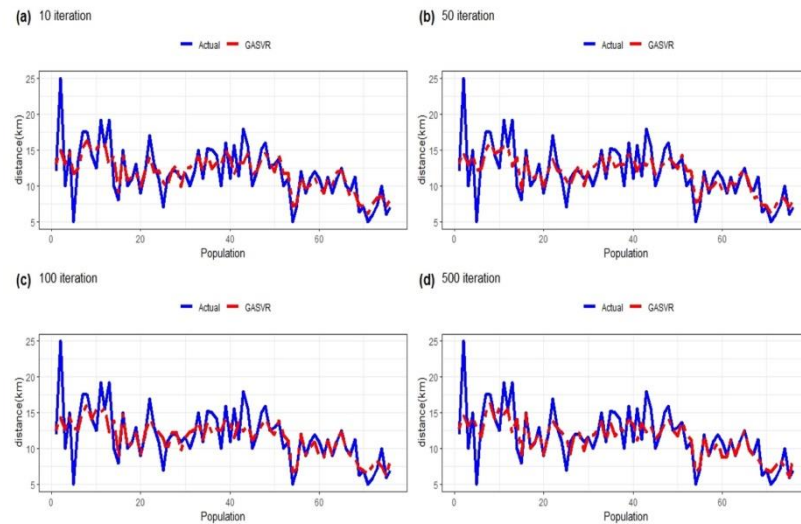


Figure 4: Plots between actual and forecast data sets

References

- [1] Buttar, H.S., Li, T. and Ravi, N. Prevention of cardiovascular diseases: Role of exercise, dietary interventions, obesity, and smoking cessation. *Experimental and Clinical Cardiology*. 2005. 10(4):229-249.
- [2] Cronkleton, E. *11 Benefits of Cycling, Plus Safety Tips*. Available: <https://www.healthline.com/health/fitness-exercise/cycling-benefits>, 2020.
- [3] Mandal, A. Just 30 minutes of exercise each day can lower heart disease risk. Online. Available: shorturl.at/cqvLQ.
- [4] Garber, C.E., Blissmer, B., Deschenes, M.R., Franklin, B., Lamonte, M.J., Lee, I.M., Nieman, D.C. and Swain, D.P. Quantity and Quality of Exercise for Developing and Maintaining Cardiorespiratory, Musculoskeletal, and Neuromotor Fitness in Apparently Healthy Adults. *Medicine & Science in Sports & Exercise*. 2011. 43(7): 1334-1359.
- [5] Research and Markets. *Fitness Equipment: COVID-19*. Available: <https://www.researchandmarkets.com>
- [6] Zander, A., Passmore, E., Mason, C. and Rissel, C. Joy, Exercise, Enjoyment, Getting out: A Qualitative Study of Older People's Experience of Cycling in Sydney, Australia. *Journal of Environmental and Public Health*. 2013. 2013(5):547453.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table 4: Comparisons of the efficiency of GA-SVR

Method	Time (minutes)	RMSE	MAPE
GA-SVR			
10 iteration	0.35	1.529	0.113
50 iteration	1.03	1.501	0.108
100 iteration	2.00	1.389	0.094
500 iteration	10.02	1.314	0.091

- [7] Titze, S., Geus, B.D., Krenn, P., Bauman, A., Oja, P., Nash, B.R. and Kohlberger, T. Health benefits of cycling: A systematic review. *Scandinavian Journal of Medicine and Science in Sports*. 2011. 21(4): 496-509.
- [8] Chuentawat, R., Kerdprasop, K. and Kerdprasop, N. The Demand Forecasting of Electrical Power Distribution Units by Support Vector Regression with 3 Folds Cross-Validation. *Journal of science and technology, Ubon Ratchathani University*. 2017. 19(1).
- [9] Ogcü, G., Demirel, O.F., and Zaim, S. Forecasting Electricity Consumption with Neural Networks and Support Vector Regression. *Procedia - Social and Behavioral Sciences*. 2012. 58: 1576-1585.
- [10] Xu, L., Zhang, W., Zhao, J., Cao, Z., Xie, R., Liu, X. and Hu, J. Support-vector-regression-based prediction of water holdup in horizontal oil-water flow by using a bicircular conductance probe array. *Flow Measurement and Instrumentation*. 2017. 57: 64-72.
- [11] Yangming, G., Lu, Z., Xiaobin, C., Congbao, R., Zhengjun, Z. and Jiezhong, M. Time Series Adaptive Online Prediction Method Combined with Modified LS-SVR and AG. *Mathematical Problems in Engineering*. 2012.
- [12] Suzukia, Y., Ibayashia, H., Kanedab, Y. and Minenoa, H. Proposal to sliding window-based support vector regression-KES2014. *18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*. 2014. 35: 1615-1624.
- [13] Yang, C.C. and Shieh, M.D. A support vector regression based prediction model of affective responses for product form design. *Computers & Industrial Engineering*. 2010, 59, 682-689
- [14] Zangooc, M.H., Habibi, J. and Alizadehsani, R. Disease Diagnosis with a hybrid method SVR using NSGA- II. *Neurocomputing*. 2014. 136(14-29).
- [15] Tang, X., Wang, L., Chang, J., Chen, J. and Sheng, V. S. Forecasting model based on information-granulated GA-SVR and ARIMA for producer price index. 2019.
- [16] Guo, M. L., Li, D., Du, C., Jia, Z., Qin, X., Chen, L., Sheng, L. and Li, H. Prediction of the Busy Traffic in Holidays Based on GA-SVR. *Advances in Computer Science and Information Engineering*. 2012. 577-582.
- [17] Zhou, J., Chang, H., Cheng, X. and Zhao, X. A Multiscale and High-Precision LSTM-GASVR Short-Term Traffic Flow Prediction Model. *Complexity*. 2020.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- [18] Liu, S., Tai, H., Ding, Q., Li, D., Xu, L. and Wei, Y. A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Mathematical and Computer Modelling*. 2013. 58(3–4): 458–46.
- [19] Yotharak, P., Meeklueb, P. and Kongsang, S. *Mathematics Game with Health Development*. Bachelor's Thesis. King Mongkut's Institute of Technology Ladkrabang. 2018.
- [20] Sanguansat, P. *Artificial Intelligence with Machine Learning*. Nonthaburi: IDC Premier. 2019.
- [21] Awad, M. and Khanna, R. *Support Vector Machines for Classification*. 2015.
- [22] Vapnik, V. *The Nature of Statistical Learning Theory*. New York: Springer. 1995.
- [23] Kao, Y. S., Nawata, K. and Huang, C. Y. Predicting Primary Energy Consumption Using Hybrid ARIMA and GA-SVR Based on EEMD Decomposition. *Mathematics* 2020. 2020. 8(10).
- [24] Parrella, F. Support Vector Machines for Regression. Online. Available: <http://onlinesvr.altervista.org/>.
- [25] Granata, F., Gargano, R. and Marinis, G.D. Support Vector Regression for Rainfall-Runoff Modeling in Urban Drainage: A Comparison with the EPA's Storm Water Management Model. *Water*, 2016, 8:3, 69.
- [26] Keerthi, S.S. and Lin, C.J. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*. 2003, 15(7):1667–1689.
- [27] Zhou, J., Chang, H., Cheng, X. and Zhao, X. A Multiscale and High-Precision LSTM-GASVR Short-Term Traffic Flow Prediction Model. *Complexity Problems Handled by Advanced Computer Simulation Technology in Smart Cities*. 2020.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ	นางสาวพรวิภา มีเคลือบ
วัน เดือน ปีเกิด	1 พฤษภาคม 2540
ที่อยู่ปัจจุบัน	77/8 หมู่ 3 ตำบลสิงห์ อำเภอบางระจัน จังหวัดสิงห์บุรี
ประวัติการศึกษา	2562 วิทยาศาสตรบัณฑิต สาขาคณิตศาสตร์ประยุกต์ เกรดเฉลี่ย 3.09 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง 2564 วิทยาศาสตรมหาบัณฑิต สาขาสาขาคณิตศาสตร์ประยุกต์ เกรดเฉลี่ย 3.68 สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ทุนการศึกษาที่ได้รับ	ทุนอุดหนุนการศึกษา
ผลงานทางวิชาการ	1. Forecasting A Cycling Distance based on Personal Health Data with Hybrid GA-SVR Approach. Internationl seminar on mathematics in industry (ISMI). 17-19 August 2021

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้