

ทฤษฎีจุดตรึงในปริภูมิยูคลิดเกี่ยวกับการจำแนกข้อมูล
(Fixed Point Theory in Euclidean Spaces and Data Classifications)

นางสาวกณวรรณ วงศ์พิท
รหัส 620510478

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์
มหาวิทยาลัยเชียงใหม่
ภาคการศึกษาที่ 1 ปีการศึกษา 2565

ทฤษฎีจุดตรึงในปริภูมิยูคลิดเกี่ยวกับการจำแนกข้อมูล
(Fixed Point Theory in Euclidean Spaces and Data Classifications)

กณวรรณ วงศ์ฟัก

ได้รับการพิจารณาอนุมัติให้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต
สาขาคณิตศาสตร์

คณะกรรมการควบคุมการค้นคว้าอิสระ:

..... วรณันท์ อินทก้อง ประธานกรรมการ
(รองศาสตราจารย์ ดร.วรณันท์ อินทก้อง)

..... อัญชลี เข้มเพ็ชร กรรมการ
(รองศาสตราจารย์ ดร.อัญชลี เข้มเพ็ชร)

วันที่ 19 เดือน ตุลาคม พ.ศ. 2565

ทฤษฎีจุดตรึงในปริภูมิยูคลิดเกี่ยวกับการจำแนกข้อมูล
(Fixed Point Theory in Euclidean Spaces and Data Classifications)

นางสาวกณวรรณ วงศ์ฟัก
รหัส 620510478

งานค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต
ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

กิตติกรรมประกาศ

การค้นคว้าอิสระนี้เป็นส่วนหนึ่งในการศึกษากระบวนวิชา 206499 (Independent Study) โดยมีวัตถุประสงค์เพื่อให้ นักศึกษาสาขาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ได้ศึกษาค้นคว้าอิสระในหัวข้อทางคณิตศาสตร์ ซึ่งผู้ค้นคว้าได้ศึกษาในเรื่องทฤษฎีจุดตรึงในปริภูมิยูคลิดเกี่ยวกับการจำแนกข้อมูล

ข้าพเจ้าขอขอบพระคุณ รองศาสตราจารย์ ดร.วราวุฒ อินถาก่อน ประธานกรรมการสอบในฐานะอาจารย์ที่ปรึกษาการค้นคว้าอิสระนี้ที่ได้ให้คำแนะนำและความรู้ต่างๆ เกี่ยวกับการทำงานค้นคว้าอิสระ ติดตามความก้าวหน้าในการดำเนินการ รวมถึงการนำเสนอและรูปเล่มรายงาน ตลอดจนตรวจสอบความสมบูรณ์ถูกต้องของการค้นคว้าอิสระนี้เสร็จสมบูรณ์ได้ด้วยดี

ขอขอบพระคุณ รองศาสตราจารย์ ดร.อัญชลี เข้มเพ็ชร อาจารย์กรรมการสอบงานค้นคว้าอิสระในครั้งนี้ที่ได้สละเวลาให้คำแนะนำ รวมถึงตรวจสอบความถูกต้อง ในการค้นคว้าอิสระนี้ให้มีความสำเร็จให้ดียิ่งขึ้น

ขอขอบคุณ นางสาวปนัดดา ทองแผ่น และนางสาวสุภารัตน์ ทองแผ่น รุ่นพี่ในภาควิชาที่กรุณาสอนการใช้แบบจำลองการจำแนกข้อมูลโดยใช้โปรแกรม Matlab

ทั้งนี้ขอกราบขอบพระคุณ บิดา มารดา ผู้มีพระคุณ ที่ให้กำลังใจตลอดการทำงานตั้งแต่เริ่มต้นจนสำเร็จลุล่วงตามวัตถุประสงค์

ขอขอบคุณคณาจารย์ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ ที่ได้ให้ความกรุณาอบรมสั่งสอนและมอบความรู้ทั้งใน และนอกตำราให้แก่ข้าพเจ้าอันเป็นประโยชน์และสามารถนำมาประยุกต์ใช้ในการศึกษาค้นคว้าอิสระนี้

และสุดท้ายนี้หากผลงานการค้นคว้าอิสระฉบับนี้ได้รับคำชื่นชม หรือเกิดประโยชน์ประการใด ข้าพเจ้าขอมอบคำชื่นชมและคุณความดีนั้นให้แก่ บิดา มารดา อาจารย์ที่ปรึกษา คณาจารย์ ตลอดจนผู้มีพระคุณทุกท่านที่ได้อบรมสั่งสอนข้าพเจ้า หากมีข้อผิดพลาดประการใดที่อาจเกิดขึ้น ข้าพเจ้าขอน้อมรับแต่เพียงผู้เดียวและยินดีรับฟังคำแนะนำจากทุกท่านที่ได้เข้ามาศึกษา และต้องขอภัยไว้ ณ ที่นี้

กณวรรณ วงศ์พัก

หัวข้อ (ภาษาไทย)	ทฤษฎีจุดตรึงในปริภูมิยูคลิดเกี่ยวกับการจำแนกข้อมูล
(ภาษาอังกฤษ)	Fixed Point Theory in Euclidean Spaces and Data Classifications
ชื่อผู้ทำการค้นคว้าอิสระ	นางสาวกณวรรณ วงศ์พัก รหัสนักศึกษา 620510478
ชื่ออาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.วารุณันท์ อินธาก้อน
ชื่อกรรมการสอบ	รองศาสตราจารย์ ดร.อัญชลี เข้มเพ็ชร

บทคัดย่อ

จุดประสงค์ของการค้นคว้าอิสระนี้ คือ ศึกษาทฤษฎีบทจุดตรึงที่สำคัญในปริภูมิยูคลิดจนถึงศึกษาวิธีการประมาณค่าบางวิธีเพื่อใช้ในการประมาณค่าจุดตรึงของการส่งแบบไม่ขยาย และศึกษาบทประยุกต์ของทฤษฎีจุดตรึงในการจำแนกข้อมูลผู้ป่วยโรคหัวใจ

คำสำคัญ : การส่งแบบไม่ขยาย, การจำแนกข้อมูล, ทฤษฎีบทจุดตรึง

Abstract

The purpose of this research are learning the important fixed point theorems in Euclidean spaces. Including, we have been studying some approximation methods for fixed point of nonexpansive mappings and applications of fixed point theorems to heart disease data classification.

Keywords : Nonexpansive mapping, data classification, fixed point theory

สารบัญ

กิตติกรรมประกาศ	i
บทคัดย่อ	ii
สารบัญ	iii
1 บทนำ (Introduction)	1
2 ความรู้พื้นฐาน (Preliminaries)	4
2.1 ปริภูมิยูคลิเดียน (Euclidean spaces)	4
2.2 ทฤษฎีจุดตรึง (Fixed points theory)	7
2.3 การจำแนกข้อมูล (Data classifications)	12
3 ผลการศึกษา (Results)	14
3.1 แบบจำลองที่ใช้ในการจำแนกข้อมูล	14
3.2 ทฤษฎีบทจุดตรึงกับการจำแนกข้อมูลผู้ป่วยที่เป็นโรคหัวใจ	16
3.2.1 ระเบียบวิธีการทำซ้ำแบบมานน์ (Mann Iteration)	17
3.2.2 ระเบียบวิธีการทำซ้ำแบบอิชิคาวา (Ishikawa Iteration)	19
4 สรุปผลการศึกษา (Conclusion)	22
บรรณานุกรม	24
ภาคผนวก	26

บทที่ 1

บทนำ (Introduction)

ทฤษฎีบทจุดตรึง (Fixed point theorem) ถือว่าเป็นทฤษฎีที่สำคัญมากในการนำไปประยุกต์ทั้งในสาขาคณิตศาสตร์ และในสาขาอื่นๆ การศึกษาวิจัยในเรื่องของทฤษฎีบทจุดตรึงบนปริภูมิบานาค เป็นการหาเงื่อนไขที่เพียงพอที่ทำให้การส่ง T ที่ส่งจากเซตย่อย X ของปริภูมิบานาค E ไปยัง X ที่มีจุดตรึง นั่นคือจะมีจุด x ใน X ซึ่งทำให้ $T(x) = x$

กระบวนการสร้างลำดับด้วยระเบียบวิธีการทำซ้ำ ในปริภูมิเวกเตอร์ต่างๆ และการศึกษาเงื่อนไขที่จำเป็น และเพียงพอที่จะทำให้ลำดับเหล่านั้นเข้าสู่จุดตรึงจุดหนึ่งของฟังก์ชันต่างๆ เป็นเรื่องที่ต้องคิดค้นและศึกษา เพื่อให้ขอบเขตการใช้งานกว้างขึ้นและง่ายขึ้น กระบวนการสร้างลำดับด้วยระเบียบวิธีการทำซ้ำมีหลากหลายวิธีด้วยกัน เช่น Mann Iteration process, Ishikawa Iteration process เป็นต้น ซึ่งเราสามารถนำเอาเทคโนโลยีทางการคำนวณด้วยคอมพิวเตอร์มาช่วยในการหาจุดตรึงโดยกระบวนการสร้างลำดับที่กล่าวมาแล้วข้างต้น

การศึกษาคณิตศาสตร์ทางการวิเคราะห์เชิงฟังก์ชัน (Functional Analysis) นั้น เรื่องของจุดตรึง (Fixed points) ได้รับการหยิบยกขึ้นมาศึกษาอยู่เสมอ ซึ่งพอจะแบ่งแยกออกเป็นประเด็นใหญ่ๆ ดังนี้

ประเด็นแรก คือ การหาเงื่อนไขที่เพียงพอให้กับปริภูมิต่างๆ เพื่อที่จะทำฟังก์ชันนั้นมีจุดตรึง เมื่อศึกษาการมีคำตอบของสมการต่างๆ แล้วปัญหาที่น่าสนใจต่อไปคือ

ประเด็นที่สอง เราจะหาคำตอบของสมการต่างๆ นั้นได้อย่างไร คำถามดังกล่าวทำให้มีนักคณิตศาสตร์จำนวนมากสนใจศึกษา และคิดค้นระเบียบวิธีการทำซ้ำของจุดตรึง (fixed-point iterations) ต่างๆ ที่ใช้ในการหาคำตอบโดยวิธีประมาณค่า และในยุคปัจจุบันปัญหาที่น่าสนใจในการศึกษาทฤษฎีจุดตรึง คือ การประยุกต์ใช้ในการแก้ปัญหาหาค่าน้อยสุด (minimization problem) ซึ่งปัญหาดังกล่าวเป็นปัญหาที่สำคัญที่มีประโยชน์มากมายในการจำแนกข้อมูลต่างๆ เช่น การจำแนกชุดข้อมูลโรคพาร์กินสัน การจำแนกชุดข้อมูลโรคหัวใจ การจำแนกชุดข้อมูลของดอกไอริส และการจำแนกชุดข้อมูลของหอยเปาฮื้อ เป็นต้น

ประเด็นที่สาม คือ การประยุกต์ใช้ทฤษฎีบทจุดตรึง ซึ่งหัวข้อที่น่าสนใจหนึ่ง คือ ทฤษฎีบทจุดตรึงกับการจำแนกข้อมูลผู้ป่วยที่เป็นโรคหัวใจ

จุดประสงค์ของการจำแนกข้อมูล คือ การหาค่า x ที่มีค่าน้อยที่สุด โดยวิธีการที่นิยมใช้ คือ วิธีกำลังสองน้อย

สุด (Ordinary least square : OLS) ดังต่อไปนี้

$$\hat{x} = \arg \min_x \|Ax - b\|_2^2 \quad (1.1)$$

เมื่อ $\|\cdot\|_2$ คือ นอร์ม-2 (l^2 - norm) อย่างไรก็ตามแบบจำลองนี้ยังคงมีสภาวะบกพร่อง (ill-conditions)

อีกหนึ่งวิธีที่ประสบความสำเร็จในการปรับปรุงแบบจำลอง คือ การถดถอยแบบลาสโซ (least absolute shrinkage and selection operator) หรือ LASSO ที่เสนอโดย Tibshirani ในปี 1996 ซึ่งเป็นการหาผลเฉลยของ

$$\hat{x} = \arg \min_x \{\|Ax - b\|_2^2 + \lambda \|x\|_1\} \quad (1.2)$$

เมื่อ $\|\cdot\|_1$ คือ นอร์ม-1 (l^1 - norm) วิธีลาสโซยังสามารถนำไปประยุกต์ใช้กับปัญหาการถดถอยและการกู้คืนภาพ

สำหรับการแก้ปัญหา (1.1) และ (1.2) โดยทำให้เป็นสมการทั่วไปนั่นคือ การหาค่าตัวแปรที่ทำให้เกิดค่าต่ำสุดของผลรวมของสองฟังก์ชัน

$$\hat{x} = \arg \min_x \{h(x) + g(x)\} \quad (1.3)$$

เพื่อที่จะแก้สมการที่ (1.3) สมมติเงื่อนไขต่อไปนี้เป็นจริง

1. $h : \mathbb{R}^n \rightarrow \mathbb{R}$ คือ ฟังก์ชันคอนเวกซ์แบบปรับเรียบ (smooth convex function) และเป็นฟังก์ชันที่หาอนุพันธ์ได้ พร้อมด้วยเกรเดียนต์เป็นการส่งแบบลิปชิตซ์ต่อเนื่อง (L -Lipschitz continuous) เมื่อ $L > 0$ นั่นคือ

$$\|\nabla h(x) - \nabla h(y)\| \leq L\|x - y\|, \forall x, y \in \mathbb{R}^n$$

2. $g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ คือ ฟังก์ชันคอนเวกซ์กึ่งต่อเนื่องล่างโดยแท้ (proper convex and lower semi-continuous function)

หลังจากนั้นเป็นต้นมาได้มีนักคณิตศาสตร์จำนวนมากที่ทำการศึกษาวิธีเพื่อหาเงื่อนไขที่เพียงพอที่เป็นสมบัติทางเรขาคณิตของปริภูมิบานาคเพื่อใช้พิสูจน์เกี่ยวกับทฤษฎีบทจุดตรึง ต่อมาได้มีนักคณิตศาสตร์จำนวนหนึ่งทำการศึกษาวิธีหาสมบัติทางเรขาคณิตที่เป็นเงื่อนไขเพียงพอที่จะพิสูจน์ทฤษฎีบทจุดตรึงสำหรับนัยทั่วไปของการส่งแบบไม่ขยาย นั่นคือ $\|T(x) - T(y)\| \leq \|x - y\|$ ทุก $x, y \in X$

ในปีค.ศ.1922 นั้น Banach ได้ค้นพบทฤษฎีบทที่มีชื่อเสียงซึ่งถูกเรียกว่า Banach Contraction Principle Theorem ดังนี้ กำหนดให้ (X, d) เป็นปริภูมิอิงระยะทาง และ $T : C \rightarrow C$ เป็นการส่งแบบหดตัว (contraction mapping) นั่นคือถ้า $d(Tx, Ty) \leq kd(x, y)$ สำหรับบางจำนวน $k \in (0, 1)$ แล้วจะมีจุด $x \in X$ ซึ่งเป็นจุดตรึงของฟังก์ชัน T นั่นคือ $Tx = x$

ในระยะแรกที่คิดค้น Banach พบว่า สามารถใช้ระเบียบวิธีการทำซ้ำแบบง่ายๆ ที่เรียกว่า ระเบียบทำซ้ำของพิคาร์ด (Picard iteration) ในการประมาณค่าจุดตรึงของการส่ง นั่นคือ ให้ $x_0 \in X$ นิยาม $x_{n+1} = Tx_n$ เมื่อ $n = 0, 1, 2, \dots$ จะได้ $x_n = T^n x_0$ สำหรับทุกๆ $n \in \mathbb{N}$ ในกรณีที่ $T : C \rightarrow C$ เป็นการส่งที่ไม่ขยาย ดังนั้น ระเบียบ

วิธีการทำซ้ำของพิคาร์ดไม่ลู่เข้า

สำหรับกระบวนการสร้างลำดับ เพื่อให้ลู่เข้าสู่จุดตรึงนั้น กระบวนการสร้างลำดับที่สำคัญ และมีชื่อเสียงซึ่งนักวิจัยทางคณิตศาสตร์ได้หยิบยกมาอ้างอิงในการทำวิจัยในด้านนี้เสมอๆ คือในปี ค.ศ.1953 W.R. Mann ได้สร้างลำดับ $\{x_n\}$ ในปริภูมิฮิลเบิร์ต H ดังต่อไปนี้ กล่าวคือ สำหรับ $\emptyset \neq C \subset H$ ซึ่งเป็นเซตนูนปิด นิยามลำดับ

$$\begin{cases} x_0 \in C \\ x_{n+1} = \alpha_n x_n + (1 - \alpha_n) T x_n \end{cases} \quad (1.4)$$

เมื่อ $T : C \rightarrow C$ เป็นการส่งแบบไม่ขยาย และ $\{\alpha_n\} \subset [0, 1]$ เขาได้พิสูจน์ว่า ถ้า $\{\alpha_n\}$ สอดคล้องกับคุณสมบัติที่เหมาะสมแล้ว $\{x_n\}$ จะลู่เข้าแบบอ่อนไปยังจุดตรึงของ T

และในปี ค.ศ.1974 S. Ishikawa ได้เสนอลำดับแบบใหม่ที่เป็นรูปทั่วไปมากกว่าลำดับของ Mann กล่าวคือ

$$\begin{cases} x_0 \in C \\ y_n = \beta_n T x_n + (1 - \beta_n) x_n \\ x_{n+1} = \alpha_n T y_n + (1 - \alpha_n) x_n \end{cases} \quad (1.5)$$

เมื่อ $T : C \rightarrow C$ เป็นการส่งแบบไม่ขยาย และ $\{\alpha_n\} \subset [0, 1]$ เขาได้พิสูจน์ว่า ถ้า $\{\alpha_n\}$ สอดคล้องกับคุณสมบัติที่เหมาะสมแล้ว $\{x_n\}$ จะลู่เข้าแบบเข้มไปยังจุดตรึงของ T

จากที่กล่าวมาข้างต้น การค้นคว้าอิสระในหัวข้อนี้ผู้ค้นคว้ามีความสนใจที่จะศึกษานิยามและสมบัติเบื้องต้นของปริภูมิยูคลิเดียน ทฤษฎีบทจุดตรึงของการส่งแบบไม่ขยายในปริภูมิยูคลิเดียน รวมถึงศึกษารายละเอียดการจำแนกข้อมูลว่ามีความสัมพันธ์กับทฤษฎีบทจุดตรึงอย่างไร พร้อมกับศึกษาโปรแกรม Matlab เพื่อใช้ในการจำแนกข้อมูลผู้ป่วยโรคหัวใจควบคู่กัน และสุดท้ายผู้ค้นคว้าได้ทำการทดลองนาระเบียบวิธีการทำซ้ำของมานน์ และระเบียบวิธีการทำซ้ำของอิชิคาว่ามาเพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล

บทที่ 2

ความรู้พื้นฐาน (Preliminaries)

ในบทนี้จะกล่าวถึงแนวคิด สัญกรณ์ และทฤษฎีบทพื้นฐานเบื้องต้น ซึ่งเป็นพื้นฐานสำคัญสำหรับที่จะศึกษาในบทต่อไป [2]

2.1 ปริภูมิยูคลิดีียน (Euclidean spaces)

ตลอดงานคั่นคว่ำอีสระนี้ กำหนดให้ $\mathbb{R}$ คือ เซตของจำนวนจริง $\mathbb{C}$ คือ เซตของจำนวนเชิงซ้อน และ $\mathbb{N}$ คือ เซตของจำนวนนับ

บทนิยาม 2.1 ให้ $\mathbb{R}$ เป็นเส้นจำนวนจริง และนิยาม $\mathbb{R}^n$ ดังนี้

$$\mathbb{R}^n = \{x = (a_1, a_2, \dots, a_n) : a_1, a_2, \dots, a_n \in \mathbb{R}\}$$

ถ้า $x = (a_1, a_2, \dots, a_n), y = (b_1, b_2, \dots, b_n)$ และ $\alpha \in \mathbb{R}$ แล้ว $x + y$ และ $\alpha x \in \mathbb{R}^n$ โดยที่

$$x + y = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

$$\alpha x = (\alpha a_1, \alpha a_2, \dots, \alpha a_n)$$

ทฤษฎีบท 2.2 สำหรับ $x, y, z \in \mathbb{R}^n$ และ $\alpha, \beta \in \mathbb{R}$ สมบัติ (A1) - (A10) ต่อไปนี้เป็นจริง

$$(A1) \ x + y \in \mathbb{R}^n$$

$$(A2) \ \alpha x \in \mathbb{R}^n$$

$$(A3) \ (x + y) + z = x + (y + z)$$

$$(A4) \ x + y = y + x$$

$$(A5) \ x + 0 = x \text{ เมื่อ } 0 = (0, 0, \dots, 0)$$

$$(A6) \ x + (-x) = 0 \text{ เมื่อ } -x = (-a_1, -a_2, \dots, -a_n)$$

$$(A7) (\alpha\beta)x = \alpha(\beta x)$$

$$(A8) 1x = x$$

$$(A9) \alpha(x + y) = \alpha x + \alpha y$$

$$(A10) (\alpha + \beta)x = \alpha x + \beta x$$

บทนิยาม 2.3 ให้ X เป็นเซตที่ไม่ใช่เซตว่างและฟังก์ชัน $d : X \times X \rightarrow \mathbb{R}^+$ จะเรียก d ว่า ฟังก์ชันระยะทาง (distance function) ถ้าสอดคล้องกับสมบัติต่อไปนี้

$$(M1) d(x, y) \geq 0 \text{ และ } d(x, y) = 0 \text{ ก็ต่อเมื่อ } x = y \text{ สำหรับทุก } x, y \in X$$

$$(M2) d(x, y) = d(y, x) \text{ สำหรับทุก } x, y \in X$$

$$(M3) d(x, y) \leq d(x, z) + d(z, y) \text{ สำหรับทุก } x, y, z \in X$$

ค่าของระยะทาง d ที่ (x, y) เขียนแทนด้วย $d(x, y)$ หมายถึง ระยะทางระหว่าง x และ y นอกจากนี้จะเรียก (X, d) ว่า ปริภูมิอิงระยะทาง (metric space) และเรียกสมาชิกของ X ของ จุด (point) ของปริภูมิอิงระยะทาง (X, d) ด้วย X

ตัวอย่าง 2.4 ปริภูมิยูคลิดเป็นปริภูมิอิงระยะทาง โดยที่ เมื่อ $x = (x_1, x_2, \dots, x_n)$ และ $y = (y_1, y_2, \dots, y_n)$ และถ้ากำหนดให้

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (2.1)$$

เมื่อ $n = 1$ คือ ระนาบเชิงซ้อน $\mathbb{C}$ โดยกำหนดนิยามของเมตริกซ์

$$d(x, y) = |x - y| \quad (2.2)$$

จากสมการ 2.2 จะพิสูจน์ว่าสอดคล้องกับ บทนิยาม 2.3 ดังนี้

ตัวอย่าง 2.5 พิจารณาเส้นจำนวนจริง $\mathbb{R}$ และตัวอย่าง $d(x, y) = |x - y|$ สำหรับทุก $x, y \in \mathbb{R}$ ดังนั้น $(\mathbb{R}, d)$ เป็น ปริภูมิอิงระยะทาง

พิสูจน์. เห็นได้ชัดเจนว่า $d(x, y) = |x - y| \geq 0$ และ

$$d(x, y) = 0 \iff |x - y| = 0 \iff x - y = 0 \iff x = y$$

ดังนั้น (M1) เป็นจริง

จาก $d(x, y) = |x - y| = |-1||y - x| = |y - x| = d(y, x)$

จะได้ว่า (M2) เป็นจริง นอกจากนี้ จะเห็นได้ว่า

$$d(x, y) = |x - y| \leq |x - z| + |z - y| = d(x, z) + d(z, y)$$

นั่นคือ (M3) เป็นจริง □

บทนิยาม 2.6 ให้ X เป็นปริภูมิเชิงเส้น (linear space) บนฟิลด์ (Fields) $\mathbb{C}$ จะเรียกฟังก์ชัน $\langle \cdot, \cdot \rangle : X \times X \rightarrow \mathbb{C}$ ว่า

ผลคูณภายใน (inner product) บน X ถ้าสอดคล้องกับสมบัติต่อไปนี้

(B1) $\langle x, x \rangle \geq 0$ สำหรับทุก $x \in H$ และ $\langle x, x \rangle = 0$ ก็ต่อเมื่อ $x = 0$

(B2) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ สำหรับทุก $x, y, z \in H$

(B3) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ สำหรับทุก $x, y \in H$ และ $\alpha \in \mathbb{C}$

(B4) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ สำหรับทุก $x, y \in H$

เรียก $(X, \langle \cdot, \cdot \rangle)$ ว่าเป็น **ปริภูมิผลคูณภายใน (inner product spaces)** และเรียก $\langle x, y \rangle$ ว่าผลคูณภายในของ x และ y นอกจากนี้จะเรียกปริภูมิผลคูณภายในที่มีความบริบูรณ์ (complete) ว่า **ปริภูมิฮิลเบิร์ต (Hilbert spaces)**

หมายเหตุ (1) ปริภูมิผลคูณภายใน จะเรียกว่า **ปริภูมิผลคูณภายในเชิงจริง (real inner product spaces)**

เมื่อสเกลาร์ และ $\langle x, y \rangle$ เป็นจำนวนจริง และสำหรับสมบัติ (B4) จะได้ว่า $\langle x, y \rangle = \langle y, x \rangle$

(2) จากสมบัติ (B2) - (B4) สำหรับทุก $x, y, z \in X$ และ $\alpha, \beta \in \mathbb{C}$ จะได้

$$\langle x, \alpha y + \beta z \rangle = \bar{\alpha} \langle x, y \rangle + \bar{\beta} \langle x, z \rangle$$

ตัวอย่าง 2.7 ให้ $X = \mathbb{R}^n$ จะได้ว่า $\mathbb{R}^n$ เป็นปริภูมิฮิลเบิร์ตที่มีฟังก์ชัน $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ ที่กำหนดโดย

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i \quad (2.3)$$

สำหรับทุก $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ เป็นผลคูณภายในบน $\mathbb{R}^n$

พิสูจน์. พิจารณา (B1) จะได้ว่า

$$\langle x, x \rangle = \sum_{i=1}^n x_i x_i = \sum_{i=1}^n x_i^2 \geq 0$$

และ

$$\langle x, x \rangle = 0 \iff \sum_{i=1}^n x_i x_i = \sum_{i=1}^n x_i^2 = 0 \iff x = 0$$

ดังนั้น (B1) เป็นจริง

จาก (B2) จะได้ว่า

$$\langle x + y, z \rangle = \sum_{i=1}^n (x_i + y_i) z_i = \sum_{i=1}^n (x_i z_i + y_i z_i) = \sum_{i=1}^n x_i z_i + \sum_{i=1}^n y_i z_i = \langle x, z \rangle + \langle y, z \rangle$$

นั่นคือ (B2) เป็นจริง

พิจารณา (B3) จะได้ว่า

$$\langle \alpha x, y \rangle = \sum_{i=1}^n (\alpha x_i) y_i = \sum_{i=1}^n \alpha (x_i y_i) = \alpha \sum_{i=1}^n x_i y_i = \alpha \langle x, y \rangle$$

ดังนั้น (B3) เป็นจริง

นอกจากนี้ จาก (B4) จะได้

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i = \sum_{i=1}^n \overline{y_i x_i} = \overline{\langle y, x \rangle}$$

นั่นคือ (B4) เป็นจริง □

บทนิยาม 2.8 ให้ X เป็นปริภูมิเชิงเส้นบนฟิลด์ $\mathbb{K}$ และ $x \in X$ จะเรียกฟังก์ชัน $\| \cdot \| : X \rightarrow \mathbb{K}$ ว่า **นอร์ม (norm)**

เมื่อ $\| \cdot \|$ สอดคล้องคุณสมบัติต่อไปนี้

$$(N1) \|x\| \geq 0, \forall x \in X;$$

$$(N2) \|x\| = 0 \Leftrightarrow x = 0;$$

$$(N3) \|x + y\| \leq \|x\| + \|y\|, \forall x, y \in X;$$

$$(N4) \|\alpha x\| = |\alpha| \|x\|, \forall x \in X \text{ และ } \forall \alpha \in \mathbb{K}.$$

ปริภูมิเชิงเส้น X ด้วยนอร์ม $\| \cdot \|$ เรียกว่า **ปริภูมิเชิงเส้นนอร์ม (normed linear space)** และจะเรียกปริภูมิที่มี
ความบริบูรณ์ (complete) ว่า **ปริภูมิบานาค (Banach spaces)**

ตัวอย่าง 2.9 $\mathbb{R}^n$ เป็นปริภูมิบานาค โดยที่สำหรับแต่ละ $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ กำหนด $\|x\|$ ดังนี้

$$\|x\| = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}} = \sqrt{|x_1|^2 + |x_2|^2 + \dots + |x_n|^2} \quad (2.4)$$

จะแทน $\|x\|$ ของ (2.1) ด้วยสัญกรณ์ $\|x\|_2$ และสำหรับแต่ละ $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ จะนิยามนอร์ม $\|x\|_1$ ดังนี้

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$$

2.2 ทฤษฎีจุดตรึง (Fixed points theory)

บทนิยาม 2.10 [3] ให้ T เป็นการส่งบนปริภูมิ X กล่าวคือ $T : X \rightarrow X$ เรานิยาม $x \in X$ ว่าเป็น **จุดตรึง (Fixed point)** ถ้า $T(x) = x$ และให้ $F(T)$ เป็นเซตของจุดตรึงทั้งหมดของ T นั่นคือ $F(T) = \{x \in X : T(x) =$

$x\}$

ตัวอย่าง 2.11 ให้ $T : \mathbb{R} \rightarrow \mathbb{R}$ นิยามโดย $T(x) = x^3, \forall x \in \mathbb{R}$

พิจารณา $x = 0$ จะได้ว่า $T(0) = (0)^3 = 0$

และ $x = 1$ จะได้ว่า $T(1) = (1)^3 = 1$

ดังนั้น $x = 0$ และ $x = 1$ เป็นจุดตรึงของ T

บทนิยาม 2.12 ให้ X เป็นปริภูมิโน름 (norm space) และ C เป็นเซตย่อยที่ไม่เป็นเซตว่างของ X การส่ง $T : C \rightarrow C$ จะเรียกว่า การส่งแบบไม่ขยาย (nonexpansive mapping) ถ้า $\|Tx - Ty\| \leq \|x - y\|$ สำหรับทุกๆ x, y ที่เป็นสมาชิกของ C

ตัวอย่าง 2.13 ให้ $T : \mathbb{R} \rightarrow \mathbb{R}$ นิยามโดย $T(x) = 1 - x$

ให้ $x, y \in \mathbb{R}$

พิจารณา $\|T(x) - T(y)\| = \|(1 - x) - (1 - y)\| = \|-x + y\| = \|-1\| \cdot \|x - y\| = \|x - y\|$

ดังนั้น T เป็นการส่งแบบไม่ขยาย

ตัวอย่าง 2.14 ให้ $T : \mathbb{R} \rightarrow \mathbb{R}$ นิยามโดย $T(x) = x^2, \forall x \in \mathbb{R}$

จากนิยาม $\|Tx - Ty\| \leq \|x - y\| \forall x, y \in \mathbb{R}$

$$\iff \|x^2 - y^2\| \leq \|x - y\|$$

$$\iff \|(x + y)(x - y)\| \leq \|x - y\|$$

$$\iff |x + y||x - y| \leq |x - y|$$

$$\iff |x + y| \leq 1$$

ถ้า $x = 2$ และ $y = 3$ จะเห็นว่า $x \neq y$

และ $|x + y| = |2 + 3| = 5 > 1$ ขัดแย้งกัน

ดังนั้น T ไม่เป็นการส่งแบบไม่ขยาย

บทนิยาม 2.15 ให้ (X, d) เป็นปริภูมิอิงระยะทาง และ $T : X \rightarrow X$ จะกล่าวว่า T เป็น การส่งแบบลิปชิตซ์ (Lipschitzian mapping) ถ้ามี $k \geq 0$

$$d(Tx, Ty) \leq kd(x, y)$$

สำหรับทุกๆ $x, y \in X$ และเรียกค่า k ที่น้อยที่สุดว่า ค่าคงตัวลิปชิตซ์ (Lipschitz constant)

ข้อสังเกต ถ้า $K > 1$ แล้วการส่ง T จะเรียกว่าเป็น การส่งแบบหดตัว (contraction mapping)

ตัวอย่าง 2.16 ให้ $X = \mathbb{R}$ และการส่ง $T : \mathbb{R} \rightarrow \mathbb{R}$ กำหนดโดย

$$Tx = \frac{1}{2}x + 1$$

สำหรับทุก $x \in \mathbb{R}$ ดังนั้น จะได้ T เป็นการส่งหดตัวและมีจุดตรึง $x = \frac{1}{2}$ เพียงจุดเดียว

ตัวอย่างต่อไปนี้จะแสดงให้เห็นว่า มีการส่งที่ไม่เป็นการส่งหดตัว แต่การส่งนั้นมีจุดตรึงเพียงจุดเดียว

ตัวอย่าง 2.17 ให้ $X = [0, 1]$ และการส่ง $T : [0, 1] \rightarrow [0, 1]$ กำหนดโดย

$$Tx = 1 - x$$

สำหรับทุก $x \in [0, 1]$ ดังนั้น T มีจุดตรึงที่จุด 1 เพียงจุดเดียว แต่ T ไม่เป็นการส่งหดตัว

บทนิยาม 2.18 การส่ง T จะถูกเรียกว่าเป็น การส่งลิปชิตซ์เขียนแบบแอล (L-Lipschitzian) ถ้ามีค่าคงที่ $L > 0$

ซึ่งทำให้ $\|Tx - Ty\| \leq L\|x - y\|$ สำหรับทุกๆ x, y ที่เป็นสมาชิกของ C และทุกๆ $n \geq 1$

ตัวอย่าง 2.19 ให้ $T : [-1, 1] \rightarrow H$ นิยามโดย $Tx = 2|x|$ สำหรับทุกๆ $x \in [-1, 1]$

ให้ $x, y \in [-1, 1]$

$$\text{พิจารณา } |Tx - Ty| = |2|x| - 2|y|| \leq 2|x - y|$$

ดังนั้น T เป็นการส่งแบบลิปชิตซ์ โดยที่ $L = 2$

ทฤษฎีบท 2.20 หลักการส่งแบบหดตัวของบานาค

ถ้า (X, d) เป็นปริภูมิเมตริกบริบูรณ์ และ $T : X \rightarrow X$ เป็นการส่งแบบหดตัว แล้ว T จะมีจุดตรึงเพียงจุดเดียวเท่านั้น

และสำหรับ $x \in X$ จะได้ว่า $\lim_{n \rightarrow \infty} T^n x = x_0$

ข้อสังเกต ถ้า $K > 1$ แล้วการส่ง T จะเรียกว่าเป็น การส่งแบบหดตัว (contraction mapping) ก็ต่อเมื่อ มี $K > 1$

ที่ทำให้ $d(Tx, Ty) \leq kd(x, y)$ สำหรับทุกๆ $x, y \in X$

ตัวอย่าง 2.21 ให้ $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ นิยามโดย $f(x) = \frac{1}{2}x$ สำหรับทุกๆ $x \in \mathbb{R}^n$

ให้ $x, y \in \mathbb{R}^n$

$$\text{พิจารณา } \|f(x) - f(y)\|_2 = \|\frac{1}{2}x - \frac{1}{2}y\|_2 = \frac{1}{2}\|x - y\|_2$$

ดังนั้น f เป็นการส่งแบบหดตัว $\frac{1}{2}$

ทฤษฎีบท 2.22 ทฤษฎีบทจุดตรึงของบราวเวอร์

ถ้า C เป็นเซตย่อยคอนเวกซ์กะทัดรัด (convex compact subset) ที่ไม่ใช่เซตว่างของ $\mathbb{R}^n$ และ $T : C \rightarrow C$ เป็นการส่งต่อเนื่อง (continuous mapping) แล้ว T มีจุดตรึงใน C

ตัวอย่าง 2.23 ให้ B แทนบอลหนึ่งหน่วยในปริภูมิ c_0 และสำหรับแต่ละ $x = (x_1, x_2, x_3, \dots) \in B$

กำหนดให้ $T : B \rightarrow B$ นิยามโดย

$$T(x) = (1 - |x_1|, x_1, x_2, \dots)$$

จะได้ว่า T เป็นการส่งต่อเนื่อง เพราะว่า

$$\|T(x) - T(y)\| = \|x - y\|$$

สำหรับทุก $x, y \in B$ ถ้า $T(x) = x$ แล้วจะได้ $(x_1, x_2, x_3, \dots) = (1 - |x_1|, x_1, x_2, \dots)$

ดังนั้น $x_1 = x_2 = x_3 = \dots$ และจาก $x = (x_1, x_2, x_3, \dots) \in c_0$ จะได้ $\lim_{n \rightarrow \infty} x_n = 0$

นั่นคือ $x_n = 0$ สำหรับทุก $n \geq 1$ และจาก $1 - |x_1| = x_1$ จะได้ $1 = 0$ เกิดข้อขัดแย้ง

ดังนั้น $T(x) \neq x$ สำหรับทุก $x \in B$

จากตัวอย่าง 2.23 ทำให้ทราบว่าทฤษฎีบทจุดตรึงของบราวเวอร์ไม่จริงในปริภูมิบานาคที่มีมิติอนันต์ จึงทำให้เกิดเป็นสมมติฐานสำหรับพิสูจน์ทฤษฎีบทจุดตรึงของบราวเวอร์ในปริภูมิบานาคที่มีมิติอนันต์ ดังทฤษฎีบทต่อไปนี้

ทฤษฎีบท 2.24 ทฤษฎีบทจุดตรึงของชาวเดอร์

ถ้า X เป็นปริภูมิบานาคจริง (real Banach space) และ C เป็นเซตย่อยปิด คอนเวกซ์ที่มีขอบเขตและไม่เป็นเซตว่างของ X ให้ $T : C \rightarrow C$ เป็นการส่งกะทัดรัดแล้ว T มีจุดตรึงใน C

สำหรับการประมาณค่าของจุดตรึงของการส่งโดยการสร้างระเบียบวิธีการทำซ้ำแบบต่างๆ นั้น จากทฤษฎีบทหลักการส่งแบบหดตัวของบานาค โดยระยะแรกนักคณิตศาสตร์ที่คิดค้น คือ Stefan Banach พบว่า สามารถใช้ **ระเบียบวิธีทำซ้ำพิการ์ด (Picard iteration)** ในการประมาณค่าจุดตรึงของการส่ง นั่นคือ ให้ $x_0 \in X$ นิยาม $x_{n+1} = Tx_n$ เมื่อ $n = 0, 1, 2, \dots$ จะได้ $x_n = T^n x_0$ สำหรับทุกๆ $n \in \mathbb{N}$ ในกรณีที่ $T : C \rightarrow C$ เป็นการส่งที่ไม่ขยาย

ตัวอย่าง 2.25 ให้ $T : R \rightarrow R$ กำหนดโดย $Tx = 1 - x$ สำหรับทุกๆ $x \in R$

ซึ่งเห็นได้ชัดว่า T เป็นการส่งที่ไม่ขยายและมีจุดตรึง

ตัวอย่าง 2.26 ถ้าเราเลือก $x_0 = \frac{1}{3}$ พบว่า ลำดับของพิการ์ด (Picard sequence) คือ $(\frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \frac{2}{3}, \dots)$

ซึ่งหมายความว่า $(T^n x_0)$ ไม่ลู่เข้า

ดังนั้น จึงมีนักคณิตศาสตร์กลุ่มใหญ่ได้สร้างระเบียบวิธีการทำซ้ำแบบต่าง ๆ มากมาย ที่สามารถตอบคำถามที่เกี่ยวกับการประมาณค่าจุดตรึงของการส่งได้กว้างกว่าระเบียบวิธีการทำซ้ำพิการ์ด เพื่อใช้ในการประมาณค่าจุดตรึงของการส่งที่ไม่เฉพาะเป็นการส่งแบบไม่ขยาย แต่เป็นการส่งที่กว้างขวางมากขึ้น

Mann (1953) ได้สร้างระเบียบวิธีการทำซ้ำสำหรับประมาณค่าจุดตรึงของการส่งในปริภูมิฮิลเบิร์ต โดยเรียกระเบียบวิธีการทำซ้ำนี้ว่า "Mann Iteration Process"

บทนิยาม 2.27 [9] ระเบียบวิธีการทำซ้ำของมานน์ (Mann iteration process)

สำหรับการประมาณค่าจุดตรึงของการส่ง T คือ สำหรับ $n \geq 1$ ให้ $x_1 \in C$,

$$x_{n+1} = \alpha_n T x_n + (1 - \alpha_n) x_n \quad (2.5)$$

เมื่อลำดับ $\{\alpha_n\} \subset [0, 1]$

ทฤษฎีบท 2.28 [9] ให้ $T : \mathbb{R}^n \rightarrow \mathbb{R}$ เป็นการส่งแบบไม่ขยาย และ $x_1 \in \mathbb{R}^n$ ถ้า (x_n) เป็นลำดับที่สร้างจากการทำซ้ำแบบมานน์ นั่นคือ

$$x_{n+1} = \alpha_n x_n + (1 - \alpha_n) T x_n$$

และ $\sum \alpha_n = \infty$ โดยมีจำนวนจริง b ที่ $0 \leq \alpha_n \leq b < 1$ แล้ว (x_n) จะเข้าสู่จุดตรึงของ T

Ishikawa (1974) ได้สร้างระเบียบวิธีการทำซ้ำแบบใหม่สำหรับประมาณค่าจุดตรึงของการส่งในปริภูมิบานาค โดยเรียกระเบียบวิธีทำซ้ำนี้ว่า "Ishikawa Iteration Process"

บทนิยาม 2.29 [7] ระเบียบวิธีการทำซ้ำของอิชิคาว่า (Ishikawa iteration process)

อิชิคาว่าได้สร้างระเบียบวิธีการทำซ้ำแบบใหม่ที่เป็นการขยายระเบียบวิธีการทำซ้ำของมานน์ สำหรับประมาณค่าจุดตรึงของการส่ง T คือ สำหรับ $n \geq 1$ ให้

$$\begin{cases} x_1 \in C, \\ y_n = \beta_n T x_n + (1 - \beta_n) x_n, \\ x_{n+1} = \alpha_n T y_n + (1 - \alpha_n) x_n \end{cases} \quad (2.6)$$

เมื่อ $\{\alpha_n\}, \{\beta_n\} \subset [0, 1]$

ทฤษฎีบท 2.30 [7] ให้ $T : \mathbb{R}^n \rightarrow \mathbb{R}$ เป็นการส่งแบบไม่ขยาย และ $x_1 \in \mathbb{R}^n$ และให้ (x_n) เป็นลำดับที่สร้างจากการทำซ้ำแบบอิซิกาวา นั่นคือ

$$\begin{cases} y_n = \alpha_n x_n + (1 - \alpha_n) T x_n, \\ x_{n+1} = \beta_n x_n + (1 - \beta_n) T y_n \end{cases} \quad (2.7)$$

และ $\sum \alpha_n = \infty$ โดยที่ $0 \leq \alpha_n \leq b < 1$ และ $\lim_{n \rightarrow \infty} \beta_n = 0$ แล้ว (x_n) จะเข้าสู่จุดตรึงของ T

ระเบียบวิธีการทำซ้ำของมานน์ (Mann iteration process) และระเบียบวิธีการทำซ้ำของอิซิกาวา (Ishikawa iteration process) มีชื่อเสียงอย่างมาก และได้ถูกนำไปศึกษาและประยุกต์ใช้งานกันอย่างแพร่หลาย

2.3 การจำแนกข้อมูล (Data classifications)

การเรียนรู้ของเครื่องแบบสุดขีด (Extreme learning machine, ELM) [6] เป็นซึ่งเป็นอัลกอริทึม (algorithm) ที่ได้รับการศึกษาอย่างกว้างขวางในหัวข้อการวิจัยที่หลากหลายสำหรับการเรียนรู้ของเครื่องจักร (machine learning) และปัญญาประดิษฐ์ (artificial intelligence) เช่น การจำแนกใบหน้า การแบ่งส่วนภาพ การถดถอย และการจำแนกข้อมูลโดยที่ ELM ได้รับการพิสูจน์ในทางทฤษฎีว่ามีความเร็วในการเรียนรู้ที่เร็วมาก และประสิทธิภาพที่ดี ซึ่ง ELM ได้ถูกนำไปใช้ในการแก้ปัญหาค่าน้อยสุด (minimization problem) เรียกว่า วิธีกำลังสองน้อยสุด (Ordinary least square, OLS)

$$\min_{\beta} \|H\beta - T\|_2^2 \quad (2.8)$$

โดยที่ $\|\cdot\|_2$ คือ l_2 - นอร์ม กำหนดโดย $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$

$T \in \mathbb{R}^{N \times m}$ เป็นเป้าหมายของข้อมูล

$\beta \in \mathbb{R}^{M \times m}$ คือ น้ำหนักที่เชื่อมระหว่างชั้นซ่อน (hidden layer) และชั้นนำออก (output layer)

$H \in \mathbb{R}^{N \times M}$ คือ เมทริกนำออกของชั้นซ่อน

M = จำนวนชุดข้อมูลที่จะนำมาทดสอบ (Test data set)

N = จำนวนชุดข้อมูลที่จะนำมาให้เครื่องเรียนรู้ (Traning data set)

เป้าหมายของแบบจำลอง คือ การหาค่าของพารามิเตอร์ β โดยการใช้วิธีการประมาณค่า

ในความเป็นจริงจำนวนของตัวแปร M ที่ไม่ทราบค่าจะมีค่ามากกว่าจำนวนข้อมูล N ของการฝึกฝน (Training data N) ซึ่งทำให้โครงข่ายเกิดปัญหา Overfitting ดังนั้น เพื่อหลีกเลี่ยงปัญหานี้ จึงเลือกใช้ การถดถอยแบบลาสโซ (Least absolute shrinkage and selection operator, LASSO)

$$\min_{\beta} \|H\beta - T\|_2^2 + \lambda \|\beta\|_1 \quad (2.9)$$

โดยที่ $\|\cdot\|_1$ - นอร์ม กำหนดโดย $\|x\|_1 = \sum_{i=1}^n |x_i|$ และ $\lambda > 0$ เรียกว่า regularization parameter

ในการหาค่าพารามิเตอร์ β โดยกำหนดให้ $f(x) = \|H\beta - T\|_2^2$ และ $g(\beta) = \lambda\|\beta\|_1$ เมื่อ $x \in \mathbb{R}^n$ และสามารถเขียนสมการ (2.6) ใหม่ได้ในรูป

$$\min_{\beta} \{f(\beta) + g(\beta)\} \quad (2.10)$$

นั่นคือ $prox_{\lambda g}\beta = \bar{x}$ ก็ต่อเมื่อ $g(\bar{x}) + \frac{1}{2\lambda}\|\bar{x} - \beta\|^2$ มีค่าน้อยที่สุด

บทนิยาม 2.31 [5] ให้ g เป็นการส่งคอนเวกซ์ และ $\lambda > 0$ ตัวดำเนินการพรอกซิมอลของพารามิเตอร์ λ ของฟังก์ชัน

g (proximity operator of parameter λ of g) ที่จุด $\beta \in H$ แทนด้วย $prox_{\lambda g}$ นิยามโดย

$$prox_{\lambda g}\beta = \underset{y \in H}{argmin} \{g(y) + \frac{1}{2\lambda}\|y - \beta\|^2\}$$

และได้มีการพิสูจน์ใน [10] ว่า β จะเป็นคำตอบของปัญหา (2.5) ค่าน้อยที่สุด ก็ต่อเมื่อ

$$\beta = prox_{cg}(I - c\nabla f)(\beta) \quad (2.11)$$

โดยที่ $c > 0$

I คือ ตัวดำเนินการเอกลักษณ์

จาก (2.8) สามารถเขียนสมการจุดตรึงได้ดังนี้

$$\beta = T(\beta) \quad (2.12)$$

โดยที่ $T = prox_{cg}(I - c\nabla f)$

บทที่ 3

ผลการศึกษา (Results)

3.1 แบบจำลองที่ใช้ในการจำแนกข้อมูล

แบบจำลองที่ใช้ในการจำแนกข้อมูลที่ศึกษาในหัวข้อนี้ [12] คือ

$$H\beta = O \quad (3.1)$$

โดยที่

$$\mathbf{H} = \begin{bmatrix} G(w_1x_1 + b_1) & \dots & G(w_Lx_1 + b_L) \\ \vdots & \ddots & \vdots \\ G(w_1x_P + b_1) & \dots & G(w_Lx_P + b_L) \end{bmatrix}_{P \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \beta_2^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}$$

และ

$$\mathbf{O} = \begin{bmatrix} o_1^T \\ o_2^T \\ \vdots \\ o_P^T \end{bmatrix}_{P \times m}$$

เมื่อ

H คือ ข้อมูลของผู้ป่วยโรคหัวใจ

O คือ ชุดข้อมูลนำออกเพื่อที่จะนำข้อมูลไปทำนายในส่วนของ Training data set และ Test data set

N คือ จำนวนชุดตัวอย่าง

G คือ ฟังก์ชันซิกมอยด์ ($\sigma(x) = \frac{1}{1+e^{(-x)}}$)

จาก

$$H\beta = O$$

ก็ต่อเมื่อ

$$\|H\beta - O\|_2^2 = 0$$

ดังนั้น เป้าหมาย คือ หา β ที่ทำให้ $\|H\beta - O\|_2^2$ มีค่าน้อยสุด

แก้ปัญหาดังกล่าวด้วยวิธีปัญหาค่าน้อยที่สุด (minimization problem) โดยการหาค่าน้ำหนักที่น้อยที่สุด ผ่านวิธีกำลังสองน้อยสุด (Least squares : LS) ดังต่อไปนี้

$$\min_{\beta} \|H\beta - O\|_2^2 \quad (3.2)$$

ซึ่งวิธีกำลังสองน้อยสุดมักจะเกิด Overfitting หมายถึง โมเดลที่ได้จาก Training data set มีความถูกต้อง แต่เมื่อนำไปใช้กับ Test data set ได้ค่าความถูกต้องต่ำ หรือตัวแบบที่ได้เป็นการเรียนรู้ข้อมูลจาก Training data set ดีมาก แต่ไม่สามารถนำไปใช้กับข้อมูลที่ไม่เคยพบมาก่อนได้ดี

เพื่อป้องกัน Overfitting จึงเลือกใช้วิธี Least absolute shrinkage and selection operator (LASSO) ดังนี้

$$\min_{\beta} \|H\beta - O\|_2^2 + \lambda \|\beta\|_1 \quad (3.3)$$

โดยที่ λ คือ regularization parameter และ กำหนดให้ $f(\beta) = \|H\beta - O\|_2^2$ และ $g(\beta) = \lambda \|\beta\|_1$
การแก้ปัญหาดังกล่าวด้วยวิธี LASSO โดยทำให้เป็นสมการทั่วไป คือการหาค่าตัวแปรที่ทำให้เกิดค่าต่ำสุดของผลรวมของทั้งสองฟังก์ชัน

$$\min_{\beta} \{f(\beta) + g(\beta)\} \quad (3.4)$$

β คือ ค่าน้อยสุดของค่า $f + g$ ก็ต่อเมื่อ [10]

$$\beta = \text{prox}_{cg}(I - c\nabla f)(\beta) \quad (3.5)$$

จากบทนิยาม 2.33 $c > 0$ ตัวดำเนินการพรอกซิมอลของพารามิเตอร์ c ของฟังก์ชัน g

ทำให้ได้ทฤษฎีบทที่ถูกต้อง ดังต่อไปนี้

$$\beta = T(\beta) \quad (3.6)$$

โดยที่ $T = \text{prox}_{cg}(I - c\nabla f)$ เรียกว่า ตัวดำเนินการไปข้างหน้า-ย้อนกลับ (forward-backward operator)

ทฤษฎีบท 3.1 [5] กำหนดให้ $f(\beta) = \|H\beta - O\|_2^2$ และ $g(\beta) = \lambda \|\beta\|_1$ จะได้ว่า $\text{prox}_{cg}(I - c\nabla f)$ เป็นการส่ง

แบบไม่ขยาย (nonexpansive mapping)

ในหัวข้อนี้จะศึกษาการจำแนกข้อมูลของผู้ป่วยโรคหัวใจจากกลุ่มตัวอย่าง 303 ตัวอย่าง ซึ่งนำข้อมูลมาจาก <https://archive.ics.uci.edu>, accessed on.

ชุดข้อมูลผู้ป่วยโรคหัวใจ แบ่งได้เป็น 2 classes คือ ต้องการจำแนกข้อมูลที่ได้รับมาว่าประชากรเป็นผู้ป่วยโรคหัวใจหรือไม่ และแบ่งข้อมูลลักษณะเฉพาะส่วนบุคคลได้ 13 attributes ได้แก่ อายุ เพศ และข้อมูลต่างๆ เกี่ยวกับการแพทย์อีก 11 attributes

ในการทดลอง เราใช้ฟังก์ชันซิกมอยด์ (sigmoid function) เหมือนกับฟังก์ชันการกระตุ้น (activation function) กำหนดให้ $\lambda = 1$ สำหรับข้อมูลของผู้ป่วยโรคหัวใจ จำนวนของ hidden nodes $L = 30$ และแบ่งชุดข้อมูลตัวอย่างที่จะนำไปใช้ในการ train 210 ข้อมูล คิดเป็นร้อยละ 70 ของข้อมูลทั้งหมด และแบ่งชุดข้อมูลตัวอย่างที่จะนำไปใช้ในการ test 93 ข้อมูล คิดเป็นร้อยละ 30 ของข้อมูลทั้งหมด รวมตัวอย่างข้อมูลที่ใช้ในการจำแนกข้อมูลทั้งสิ้น 303 ตัวอย่าง สำหรับการจำแนกข้อมูลในครั้งนี้

อัลกอริทึม ELM

กำหนดให้ $S = \{(x_j, o_j) : x_j \in \mathbb{R}^n, o_j \in \mathbb{R}^m, j = 1, 2, \dots, P\}$ เป็นเซตของ Training

โดยที่ $x_j = [x_{j1}, x_{j2}, \dots, x_{jn}]^T \in \mathbb{R}^n$ คือ ชุดข้อมูลนำเข้า (input data)

และ $o_j = [o_{j1}, o_{j2}, \dots, o_{jm}]^T \in \mathbb{R}^m$ คือ ผลลัพธ์ของ o_j

ขั้นตอนที่ 1 : เลือก regularization parameter λ และจำนวน M ใน hidden node

ขั้นตอนที่ 2 : สุ่ม w_i และ b_i เมื่อ $i = 1, 2, \dots, M$

ขั้นตอนที่ 3 : คำนวณ hidden layer output matrix H

ขั้นตอนที่ 4 : คำนวณ β โดยใช้ $H\beta = O$

ผลลัพธ์ของ Extreme Learning Machine (ELM) สำหรับ Single hidden layer feedforward networks (SLFN) ด้วย L ของ hidden node จะได้ว่า

$$o_j = \sum_{i=1}^L \beta_i G(\langle w_i, x_j \rangle + b_i)$$

เมื่อ $w_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in \mathbb{R}^n$ คือ เวกเตอร์น้ำหนักที่เชื่อม hidden node ที่ i และ input node

$\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T \in \mathbb{R}^m$ คือ ผลลัพธ์ของน้ำหนัก (output weight) ของ hidden node ที่ i

b_i หรือ bias คือ จำนวนที่บวกเข้าไปเพื่อปรับค่า ให้ค่าที่คำนวณออกมาถูกต้องมากขึ้น

3.2 ทฤษฎีบทจุดตรึงกับการจำแนกข้อมูลผู้ป่วยที่เป็นโรคหัวใจ

ในหัวข้อนี้จะศึกษาวิธีการนำระเบียบวิธีในทฤษฎีบทจุดตรึงของการส่งแบบไม่ขยายต่างๆ ไปในการจำแนกข้อมูลผู้ป่วยที่เป็นโรคหัวใจ โดยระเบียบวิธีที่นำมาศึกษาในหัวข้อนี้ ได้แก่

1. ระเบียบวิธีการทำซ้ำแบบมานน์ (Mann Iteration)
2. ระเบียบวิธีการทำซ้ำแบบอิชิคาว่า (Ishikawa Iteration)

การจำแนกข้อมูลสามารถจำแนกได้ ดังนี้

กำหนดให้ **True Positive (TP)** คือ จำนวนผู้ป่วยเป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าเป็นผู้ป่วยโรคหัวใจ

False Negative (FN) คือ จำนวนผู้ป่วยเป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

True Negative (TN) คือ จำนวนประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

False Positive (FP) คือ จำนวนประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นผู้ป่วยโรคหัวใจ

ค่า Pre คือ ทำนายคนเป็นโรคหัวใจว่าเป็นผู้ป่วยโรคหัวใจ

ค่า Rec คือ ทำนายคนเป็นโรคหัวใจว่าไม่เป็นผู้ป่วยโรคหัวใจ หรือสุขภาพดี

Accuracy คือ ค่าความถูกต้อง สามารถคำนวณได้จาก

$$Accuracy = \frac{\text{correct predicted data}}{\text{all data}} \times 100$$

3.2.1. ระเบียบวิธีการทำซ้ำแบบมานน์ (Mann Iteration)

กำหนดให้ $T(x) = \text{prox}_{cg}(I - c\nabla f)(x)$

โดยที่ $f(x) = \|Hx - O\|_2^2$, $g(x) = \lambda\|x\|_1$ และ $\alpha_k = 0.5$ โดยใช้ระเบียบวิธีทำซ้ำแบบมานน์

$$x_{n+1} = \alpha_k T x_k + (1 - \alpha_k) x_k \quad (3.7)$$

จากทฤษฎีบท 3.1 จะได้ว่า T เป็นการส่งแบบไม่ขยาย และจากทฤษฎีบท 2.28 จะได้ว่า T มีจุดตรึงซึ่งในที่นี้คือ ค่า β ที่ต้องการในปัญหาสมการ 3.6 และส่วนของ $T x_k$ ในโปรแกรมกำหนดให้ $T(x_0, \text{lamda}_k, \text{para.lam}, A, b)$ x_0 คือ เมทริกซ์เริ่มต้นที่มี ขนาด $m \times n$

$\text{lamda}_k = \frac{i}{(L_F * (i+1))}$ ซึ่งค่า $i = 1 : 50$ จะอยู่ใน Loop for หมายถึง ทำตั้งแต่ 1 ถึง 50 และ L_F คือ ค่าของลิฟชิตซ์ $\text{para.lam} = \text{lam}$ คือ 1×10^{-5}

A คือ ชุดข้อมูลนำเข้าโรคหัวใจมีทั้งผู้ป่วยเป็นโรคหัวใจ และไม่ได้ป่วยเป็นโรคหัวใจที่ยังไม่ได้จำแนก 303 ตัวอย่าง

b คือ ชุดข้อมูลที่ชื่อว่า $data$ ซึ่งใน $data$ มีชุดข้อมูลย่อย 4 ชนิดเพื่อใช้ในการจำแนกข้อมูล ดังนี้

1. $data.xTrain$ คือ ชุดข้อมูลผู้ป่วย 210 ตัวอย่าง และข้อมูลลักษณะเฉพาะส่วนบุคคล 13 attributes
2. $data.tTrain$ คือ ผลลัพธ์ของผู้ป่วยที่จะนำไปทำนาย

ผลลัพธ์หมายเลข 1 หมายถึง ประชากรสุขภาพดีและไม่ได้เป็นโรคหัวใจ มีจำนวน 95 คน จากทั้งหมด 210 คน

และผลลัพธ์หมายเลข 2 หมายถึง ผู้ป่วยเป็นโรคหัวใจ มีจำนวน 115 คน จากทั้งหมด 210 คน

3. $data.xTest$ คือ ชุดข้อมูลผู้ป่วย 93 ตัวอย่าง และข้อมูลลักษณะเฉพาะส่วนบุคคล 13 attributes
4. $data.tTest$ คือผลลัพธ์ของผู้ป่วยที่จะนำไปทำนาย

หมายเลข 1 หมายถึง ประชากรสุขภาพดีและไม่ได้เป็นโรคหัวใจ มีจำนวน 43 คน จากทั้งหมด 93 คน

และหมายเลข 2 หมายถึง ผู้ป่วยเป็นโรคหัวใจ มีจำนวน 50 คน จากทั้งหมด 93 คน

ได้ผลลัพธ์ ดังนี้

TP(train), TP(test) = 102.00 , 41.00

ค่า TP(train) คือ จำนวน 102 คนจาก 210 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ และทำนายถูกว่าเป็นโรคหัวใจ

ค่า TP(test) คือ จำนวน 41 คนจาก 93 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ และทำนายถูกว่าเป็นโรคหัวใจ

FN(train), FN(test) = 13.00, 9.00

ค่า FN(train) คือ จำนวน 13 คนจาก 210 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็น ประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

ค่า FN(test) คือ จำนวน 9 คนจาก 93 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็น ประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

TN(train), TN(test) = 54.00, 27.00

ค่า TN(train) คือ จำนวน 54 คนจาก 210 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

ค่า TN(test) คือ จำนวน 27 คนจาก 93 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

FP(train), FP(test) = 41.00,16.00

ค่า FP(train) คือ จำนวน 41 คนจาก 210 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นผู้ป่วยโรคหัวใจ

ค่า FP(test) จำนวน 16 คนจาก 93 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นผู้ป่วยโรคหัวใจ

Pre train = 0.7133

ค่า Pre train คือ ทำนายคนเป็นโรคหัวใจว่า เป็นผู้ป่วยโรคหัวใจ ในส่วนของ Training ได้ร้อยละ 71.33

Rec train = 0.8870

ค่า Rec train คือ ทำนายคนเป็นโรคหัวใจว่า ไม่เป็นผู้ป่วยโรคหัวใจ หรือสุขภาพดี ในส่วนของ Training ได้ร้อยละ 88.70

Pre test = 0.7193

ค่า Pre test คือ ทำนายคนเป็นโรคหัวใจว่า เป็นผู้ป่วยโรคหัวใจ ในส่วนของ Test ได้ร้อยละ 71.93

Rec test = 0.8200

ค่า Rec test คือ ทำนายคนเป็นโรคหัวใจว่า ไม่เป็นผู้ป่วยโรคหัวใจ หรือสุขภาพดี ในส่วนของ Test ได้ร้อยละ 82.00

Accuracy(train, test) = 74.2857, 73.1183

ค่าความถูกต้องของ Train ได้ 74.2857

ค่าความถูกต้องของ Test ได้ 73.1183

Accuracy(train, test) = 74.29, 73.12

ค่าความถูกต้องของ Train ได้ 74.29

ค่าความถูกต้องของ Test ได้ 73.12

3.2.2. ระเบียบวิธีการทำซ้ำแบบอิชิคาวา (Ishikawa Iteration)

กำหนดให้ $T(x) = \text{prox}_{cg}(I - c\nabla f)(x)$

โดยที่ $f(x) = \|Hx - O\|_2^2$, $g(x) = \lambda\|x\|_1$, $\alpha_k = 0.5$ และ $\beta_k = 1 \times 10^{(-5)}$ โดยใช้ระเบียบวิธีการทำซ้ำแบบอิชิคาวา

$$\begin{cases} x_1 \in C \\ y_k = \beta_k T x_k + (1 - \beta_k) x_k \\ x_{k+1} = \alpha_k T y_k + (1 - \alpha_k) x_k \end{cases} \quad (3.8)$$

จากทฤษฎีบท 3.1 จะได้ว่า T เป็นการส่งแบบไม่ขยาย และจากทฤษฎีบท 2.30 จะได้ว่า T มีจุดตรึงและส่วนของ $T x_k$ ซึ่งในที่นี้ คือ ค่า β ที่ต้องการในปัญหาสมการ 3.6 ในโปรแกรม กำหนดให้ $T(x_0, \text{lamda}_k, \text{para.lam}, A, b)$

และ $T y_k$ กำหนดให้ $T(w_k, \text{lamda}_k, \text{para.lam}, A, b)$

ค่า x_0 คือ เมทริกซ์เริ่มต้นที่มี ขนาด $m \times n$

ค่า w_k คือ ค่าที่หาได้จากระเบียบวิธีการทำซ้ำ

$\text{lamda}_k = \frac{i}{(L_F * (i+1))}$ ซึ่งค่า $i = 1 : 50$ จะอยู่ใน Loop for หมายถึง ทำตั้งแต่ 1 ถึง 50 และ L_F คือ ค่าของลิฟชิตซ์
 $\text{para.lam} = \text{lam}$ คือ 1×10^{-5}

A คือ ชุดข้อมูลนำเข้าโรคหัวใจมีทั้งป่วยเป็นโรคหัวใจ และไม่ได้ป่วยเป็นโรคหัวใจที่ยังไม่ได้จำแนก 303 ตัวอย่าง

b คือ ชุดข้อมูลที่ชื่อว่า $data$ ซึ่งใน $data$ มีชุดข้อมูลย่อย 4 ชนิดเพื่อใช้ในการจำแนกข้อมูล ดังนี้

1. $data.xTrain$ คือ ชุดข้อมูลผู้ป่วย 210 ตัวอย่าง และข้อมูลลักษณะเฉพาะส่วนบุคคล 13 attributes

2. $data.tTrain$ คือ ผลลัพธ์ของผู้ป่วยที่จะนำไปทำนาย

ผลลัพธ์หมายเลข 1 หมายถึง ประชากรสุขภาพดีและไม่ได้เป็นโรคหัวใจ มีจำนวน 95 คน จากทั้งหมด 210 คน

และผลลัพธ์หมายเลข 2 หมายถึง ผู้ป่วยเป็นโรคหัวใจ มีจำนวน 115 คน จากทั้งหมด 210 คน

3. $data.xTest$ คือ ชุดข้อมูลผู้ป่วย 93 ตัวอย่าง และข้อมูลลักษณะเฉพาะส่วนบุคคล 13 attributes

4. $data.tTest$ คือผลลัพธ์ของผู้ป่วยที่จะนำไปทำนาย

หมายเลข 1 หมายถึง ประชากรสุขภาพดีและไม่ได้เป็นโรคหัวใจ มีจำนวน 43 คน จากทั้งหมด 93 คน

และหมายเลข 2 หมายถึง ผู้ป่วยเป็นโรคหัวใจ มีจำนวน 50 คน จากทั้งหมด 93 คน

ได้ผลลัพธ์ ดังนี้

TP(train), TP(test) = 103.00 , 41.00

ค่า TP(train) คือ จำนวน 103 คนจาก 210 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ และทำนายถูกว่าเป็นโรคหัวใจ

ค่า TP(test) คือ จำนวน 41 คนจาก 93 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ และทำนายถูกว่าเป็นโรคหัวใจ

FN(train), FN(test) = 12.00, 9.00

ค่า FN(train) คือ จำนวน 12 คนจาก 210 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

ค่า FN(test) คือ จำนวน 9 คนจาก 93 คน พบว่าเป็นผู้ป่วยเป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

TN(train), TN(test) = 51.00, 27.00

ค่า TN(train) คือ จำนวน 51 คนจาก 210 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

ค่า TN(test) คือ จำนวน 27 คนจาก 93 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ และแบบจำลองทำนายถูกว่าสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ

FP(train), FP(test) = 44.00, 16.00

ค่า FP(train) คือ จำนวน 44 คนจาก 210 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นผู้ป่วยโรคหัวใจ

ค่า FP(test) จำนวน 16 คนจาก 93 คน พบว่าเป็นประชากรสุขภาพดี หรือไม่ได้เป็นโรคหัวใจ แต่แบบจำลองทำนายผิดว่าเป็นผู้ป่วยโรคหัวใจ

Pre train = 0.7007

ค่า Pre train คือ ทำนายคนเป็นโรคหัวใจว่า เป็นผู้ป่วยโรคหัวใจ ในส่วนของ Training ได้ร้อยละ 70.07

Rec train = 0.8957

ค่า Rec train คือ ทำนายคนเป็นโรคหัวใจว่า ไม่เป็นผู้ป่วยโรคหัวใจ หรือสุขภาพดี ในส่วนของ Training ได้ร้อยละ 89.57

Pre test = 0.7193

ค่า Pre test คือ ทำนายคนเป็นโรคหัวใจว่า เป็นผู้ป่วยโรคหัวใจ ในส่วนของ Test ได้ร้อยละ 71.93

Rec test = 0.8200

ค่า Rec test คือ ทำนายคนเป็นโรคหัวใจว่า ไม่เป็นผู้ป่วยโรคหัวใจ หรือสุขภาพดี ในส่วนของ Test ได้ร้อยละ 82.00

Accuracy(train, test) = 73.3333, 73.1183

ค่าความถูกต้องของ Train ได้ 73.3333

ค่าความถูกต้องของ Test ได้ 73.1183

Accuracy(train, test) = 73.33, 73.12

ค่าความถูกต้องของ Train ได้ 73.33

ค่าความถูกต้องของ Test ได้ 73.12

บทที่ 4

สรุปผลการศึกษา (Conclusion)

การค้นคว้าอิสระนี้ศึกษาขั้นตอนวิธีทางทฤษฎีจุดตรึงกับการจำแนกข้อมูลผู้ป่วยโรคหัวใจ โดยใช้กระบวนการ ELM กับแบบจำลอง LASSO โดยหา β ที่ทำให้ $\|H\beta - O\|_2^2 + \lambda\|\beta\|_1$ น้อยที่สุด คือ β เป็นคำตอบของปัญหา

$$\min_{\beta} \|H\beta - O\|_2^2 + \lambda\|\beta\|_1$$

ในการหาค่า β โดยกำหนดให้ $f(\beta) = \|H\beta - O\|_2^2$ และ $g(\beta) = \lambda\|\beta\|_1$ สามารถเขียนสมการใหม่ได้ให้อยู่ในรูป

$$\min_{\beta} \{f(\beta) + g(\beta)\}$$

และได้มีการพิสูจน์ไว้แล้วว่า β จะเป็นคำตอบของปัญหาค่าน้อยที่สุด ก็ต่อเมื่อ

$$\beta = \text{prox}_{cg}(I - c\nabla f)(\beta)$$

นั่นคือ β เป็นจุดตรึงของการส่งแบบไม่ขยาย โดยที่ $c > 0$ และ I คือ ตัวดำเนินการเอกลักษณ์

สุดท้ายเราได้ใช้ระเบียบวิธีการทำซ้ำของทฤษฎีบทจุดตรึงกับการจำแนกข้อมูลผู้ป่วยที่เป็นโรคหัวใจจำนวน 303 คน ซึ่งนำข้อมูลมาจาก <https://archive.ics.uci.edu>, accessed on. และได้ผลดังนี้

ระเบียบวิธีการทำซ้ำของมานัน	Training data set	Test data set
ทำนายถูก	0.7133	0.7193
ทำนายผิด	0.8870	0.8200

ตารางที่ 4.1: ระเบียบวิธีการทำซ้ำของมานัน

ระเบียบวิธีการทำซ้ำของอิซิคาวา	Training data set	Test data set
ทำนายถูก	0.7007	0.7193
ทำนายผิด	0.8957	0.8200

ตารางที่ 4.2: ระเบียบวิธีการทำซ้ำของอิซิคาวา

สามารถสรุปผลได้ว่า ทั้ง 2 ระเบียบวิธีการทำซ้ำแบบมานน์ และระเบียบวิธีการทำซ้ำแบบอิชิตาว่า แบ่งชุดข้อมูลตัวอย่างที่จะนำไปใช้ในการ train 210 ข้อมูล คิดเป็นร้อยละ 70 ของข้อมูลทั้งหมด และแบ่งชุดข้อมูลตัวอย่างที่จะนำไปใช้ในการ test 93 ข้อมูล คิดเป็นร้อยละ 30 ของข้อมูลทั้งหมด รวมตัวอย่างข้อมูลที่ใช้ในการจำแนกข้อมูลทั้งสิ้น 303 ตัวอย่าง และจะเห็นอย่างได้ชัดว่า ระเบียบวิธีการทำซ้ำแบบมานน์และระเบียบวิธีการทำซ้ำแบบอิชิตาว่า ประสิทธิภาพยังไม่ดีพอ อย่างไรก็ตาม การศึกษาการสร้างระเบียบวิธีการทำซ้ำเป้าหมายในการวิจัยของสาขานี้ต้องการให้ค่าความถูกต้องมีค่าสูงขึ้น หรือมีค่าใกล้เคียง 100 เปอร์เซนต์

บรรณานุกรม

- [1] ภูมิ คำเอม. (2553). การประมาณค่าจุดตรึงร่วมสำหรับวงรีนัยได้ของการส่งแบบไม่ขยายและปัญหาดุลยภาพ (รายงานวิจัยฉบับสมบูรณ์). กรุงเทพฯ: มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี.
- [2] รตนกร วัฒนทวีกุล. (2564). ทฤษฎีจุดตรึงของการส่งไม่เชิงเส้น. อุบลราชธานี: มหาวิทยาลัยอุบลราชธานี.
- [3] ศรศักดิ์ เทียนหวาน. (2552). การมีจริงและการประมาณค่าวิธีทำซ้ำของจุดตรึงของการส่งที่ไม่ขยายและการส่งที่ไม่ขยายแบบเชิงเส้นกำกับ. มหาวิทยาลัยนเรศวร, 17(2), 185-190.
- [4] ศรศักดิ์ เทียนหวาน. (2553). การทำซ้ำจุดตรึงสำหรับการส่งที่ไม่ขยายและการส่งที่ไม่ขยายแบบเชิงเส้นกำกับ ในปริภูมิบานาค (รายงานวิจัยฉบับสมบูรณ์). พะเยา: สำนักวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยนเรศวร พะเยา.
- [5] Bauschke, H.H., & Combettes, L.P. (2010). Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer Nature : Cham.
- [6] Bussaban, L., Kaewkhao, A., & Suantai, S. (2020). A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. (Tech. Rep. No.1). Chiangmai, Chiang Mai University.
- [7] Ishikawa, S. (1974). Fixed points by a new iteration method. Proc. Amer. Math. Soc. 44, 147-150.
- [8] Kreyszig, E. (1978). Introductory Functional Analysis with Applications. Ontario: University of Windsor.
- [9] Mann, W.R. (1953). Mean value methods in iteration. Proc. Amer. Math. Soc. 4, 506-510.
- [10] Parikh, N., & Boyd, S. (2014). Proximal Algorithms. Found. Trends Optim, 1, 127-239.
- [11] Phuengrattana, W., & Suantai, S. (2013). Comparison of the Rate of Convergence of Various Iterative Methods for the Class of Weak Contractions in Banach Spaces, 11, 217-226.

- [12] Thongpaen, P. (2022). A New Accelerated Algorithm for Bi-Level Convex Minimization Problems and Applications in Data Prediction and Classification. Poster session presented at Mathematics Department, Faculty of Science Chiang Mai University.
- [13] Tikhonov, A.N., & Arsenin, V.Y. (1977). Solution of Ill-Posed Problems. V.H. Winston : Washington ,DC.

ภาคผนวก



Fixed Point Theory in Euclidean Spaces and Data Classifications

Author : Kanawat Wongfak ID 620510478

Advisor : Assoc.Prof.Dr.Warunun Inthakon

Department of Mathematics, Faculty of Science, Chiang Mai University.



Abstract

The purpose of this research are learning the important fixed point theorems in Euclidean spaces. Including, we have been studying some approximation methods for fixed point of nonexpansive mappings and applications of fixed point theorems to heart disease data classification.

Introduction

An element $x \in X$ is said to be a fixed point of mapping $T : X \rightarrow X$ proved $Tx = x$. The set of all fixed points of T is denoted by $F(T) = \{x \in X : Tx = x\}$.

Let X be a Banach space and C a convex subset of X . A mapping $T : C \rightarrow C$ is said to be nonexpansive if $\|Tx - Ty\| \leq \|x - y\|$ for all x, y in C .

Iteration for Data Classification Problems

In 1953, Mann has introduced The Mann iteration process is defined as follows: For C a convex subset of normed space X and nonlinear mapping T of C into itself, the sequence $\{x_n\}$ in C is defined by

$$x_1 = x \in C$$

$$x_{n+1} = \alpha_n x_n + (1 - \alpha_n)Tx_n, n \in \mathbb{N}$$

where $\{\alpha_n\}$ is real sequence in $(0,1)$.

In 1974, Ishikawa has introduced The Ishikawa iteration process is defined as follows: For C a convex subset of normed space X and nonlinear mapping T of C into itself, the sequence $\{x_n\}$ and $\{y_n\}$ in C is defined by

$$x_1 = x \in C$$

$$y_n = (1 - \beta_n)x_n + \beta_n Tx_n,$$

$$x_{n+1} = \alpha_n x_n + (1 - \alpha_n)Ty_n, n \in \mathbb{N}$$

where $\{\alpha_n\}$ and $\{\beta_n\}$ is real sequence in $(0,1)$.

Experiments for data classifications

In the experiment, we aim to classify heart disease data set from <https://archive.ics.uci.edu>, accessed on.

We obtained 303 heart disease data set that divided into 2 classes and 13 attributes which of these going to classify 210 training data set and 93 test data set.

In all experiments, we use sigmoid as an activation function and $\lambda = 10^{-5}$ for dataset with number of hidden node $L = 30$.

Applications to Data Classification Problems

The target of this model is to find the parameter β that solves the following minimization problem, called ordinary least square (OLS),

$$\min_{\beta} \|H\beta - O\|_2^2$$

where $\|\cdot\|_2$ is l_2 -norm, $O \in \mathbb{R}^{N \times m}$ is the target of data, $\beta \in \mathbb{R}^{M \times m}$ is a weight which connects hidden layer and output layer and $H \in \mathbb{R}^{N \times M}$ is the hidden layer output matrix.

In this study, we focus on the following problem, called least absolute shrinkage and selection operator (LASSO),

$$\min_{\beta} \|H\beta - O\|_2^2 + \lambda \|\beta\|_1$$

where $\|\cdot\|_1$ is l_1 -norm, $O \in \mathbb{R}^{N \times m}$ is the target of data, $\beta \in \mathbb{R}^{M \times m}$ is a weight which connects hidden layer and output layer and $H \in \mathbb{R}^{N \times M}$ is the hidden layer output matrix.

In general, LASSO can be rewritten as minimization of sum of two functions,

$$\min_{\beta} \{f(\beta) + g(\beta)\}$$

In fixed point theory, the solution of minimization can be characterized as follows: $\bar{x}$ is a minimizer of $f + g$ if and only if

$$\beta = prox_{cg}(I - c \nabla f)(\beta)$$

where $c > 0$, I is an identity operator, $prox_{cg}$ is the proximity operator of cg .

For convenience, β can be rewritten as: $\beta = T(\beta)$ where $T = prox_{cg}(I - c \nabla f)$

Results

	Training data set	Test data set
Mann iteration process		
Accurate predictions	0.7133	0.7193
Wrong predictions	0.8870	0.8200
Ishikawa iteration process		
Accurate predictions	0.7007	0.7193
Wrong predictions	0.8957	0.8200

Mann iteration process and Ishikawa iteration process are not good efficiency. However, the research of this field requires 100% accuracy.

References

- [1] Bauschke, H.H., & Combettes, L.P. (2010). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Nature : Cham.
- [2] Bunsaban, L., Kaewkhao, A., & Suanrit, S. (2020). A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. *Tech. Rep. No.1*, Chiangmai, Chiang Mai University.
- [3] Ishikawa, S. (1974). Fixed points by a new iteration method. *Proc. Amer. Math. Soc.* 44, 147-150.
- [4] Mann, W.R. (1953). Mean value methods in iteration. *Proc. Amer. Math. Soc.* 4, 506-510.
- [5] Parikh, N., & Boyd, S. (2014). *Proximal Algorithms*. *Found. Trends Optim.* 1, 127-208.
- [6] Thongsoen, P. (2023). A New Accelerated Algorithm for Bi-Level Convex Minimization Problems and Applications in Data Prediction and Classification. Poster session presented at Mathematics Department, Faculty of Science Chiang Mai University.

รูปที่ 5.1: โปสเตอร์