

โครงข่ายประสาทเทียมสำหรับการวิเคราะห์การถดถอยเชิงเส้นตามบริบทนี้ทั่วไป



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2564

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

NEURAL NETWORKS FOR GENERALIZED CONTEXTUAL LINEAR REGRESSION

Mr. Chayanon Kattiyapirak



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Statistics Program in Statistics

Department of Statistics

Faculty of Commerce and Accountancy

Chulalongkorn University

Academic Year 2021

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์

โครงข่ายประสาทเทียมสำหรับการวิเคราะห์การถดถอย
เชิงเส้นตามบริบทภัยพิบัติ

โดย

นายชยานนท์ ชัตติยาภิรักษ์

สาขาวิชา

สถิติ

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบริหารธุรกิจ



คณะกรรมการสอบวิทยานิพนธ์

(รองศาสตราจารย์ ดร. วิเลิศ ภูริวัชร)

คณบดีคณะพาณิชยศาสตร์และ
การบัญชี

ประธาน

(ผู้ช่วยศาสตราจารย์ ดร. นันท กุลวานิช)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์)

กรรมการ

(อาจารย์ ดร. สาวิตรี บุญพัชรนนท์)

กรรมการภายนอกมหาวิทยาลัย

(Professor Anthony Hayter)

ชยานนท์ ชัตติยาภิรักษ์: โครงข่ายประสาทเทียมสำหรับการวิเคราะห์การถดถอยเชิงเส้นตามบริบททั่วไป. (NEURAL NETWORKS FOR GENERALIZED CONTEXTUAL LINEAR REGRESSION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : รศ. ดร. เสกสรร เกียรติสุไพบูลย์, 60 หน้า.

ปัญหาความสัมพันธ์เชิงเส้นตามบริบท คือปัญหาที่มีตัวแปรต้นที่แบ่งข้อมูลออกเป็นกลุ่มต่าง ๆ โดยในแต่ละกลุ่มจะมีความสัมพันธ์กับผลเฉลยในลักษณะเชิงเส้นที่แตกต่างกัน ทางผู้วิจัยได้สนใจที่จะนำวิธีโครงข่ายประสาทเทียม (Neural Networks) มาแก้ไขปัญหาประเภทดังกล่าว โดยพัฒนาโครงสร้างที่ชื่อว่า Generalized Contextual Regression (GCR) และเปรียบเทียบกับโครงสร้างที่เคยมีมาก่อน ได้แก่ Feedforward Neural Networks (FNN) ซึ่งเป็นโครงสร้างพื้นฐาน และ Contextual Regression (CR) ซึ่งนำเสนอโดย Liu และ Wang (2017) งานวิจัยนี้จะศึกษาเฉพาะปัญหาการถดถอยเชิงเส้น ที่ตัวแปรต้นไม่เกิน 10 ตัว ซึ่งมีตัวแปรเชิงบริบทไม่เกิน 3 ตัวเท่านั้น โดยจากผลการวิจัยพบว่าวิธี GCR มีประสิทธิภาพสูงที่สุดในการแก้ไขปัญหาความสัมพันธ์เชิงเส้นตามบริบทเมื่อเปรียบเทียบกับวิธี FNN และ CR



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาควิชา	สถิติ	ลายมือชื่อนิสิต
สาขาวิชา	สถิติ	ลายมือชื่อ อ.ที่ปรึกษาหลัก
ปีการศึกษา	2564		

6380063526: MAJOR STATISTICS

KEYWORDS: NEURAL NETWORK / LINEAR REGRESSION / CONTEXTUAL REGRESSION

CHAYANON KATTIYAPIRAK : NEURAL NETWORKS FOR GENERALIZED CONTEXTUAL LINEAR REGRESSION. ADVISOR : ASSOC. PROF. SEKSAN KIATSUPAIBUL, Ph.D., 60 pp.

A problem with group inputs that partition observations into groups, where each group has a unique linear relationship between the response variable and explanatory variable, is known as a linear contextual relation problem. We propose to apply neural network structure to solve the problem by implementing our new method called generalized contextual regression (GCR). We also compare the model's effectiveness to the standard approaches, feedforward neural network (FNN), and the contextual regression method (GCR) by Liu and Wang (2017), which is our motivation. The problem in this study is only linear regression problems which contain at most 10 explanatory variables and at most 3 contextual explanatory variables. From the study, we find that GCR performs best in solving the contextual linear relation problem, comparing with the alternativesm, FNN and CR.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Department:	Statistics	Student's Signature
Field of Study:	Statistics	Advisor's Signature
Academic Year:	2021	

กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงได้จากการดูแล และให้ความช่วยเหลือจากรองศาสตราจารย์ ดร. เสกสรร เกียรติสุไพบูลย์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา ข้อเสนอแนะที่เป็นประโยชน์ ตลอดจนแก้ไขข้อผิดพลาดต่าง ๆ นำมาซึ่งงานวิจัยฉบับที่สมบูรณ์นี้ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ผู้วิจัยขอกราบขอบพระคุณผู้ช่วยศาสตราจารย์ ดร. นันท กุลวานิช ประธานกรรมการสอบวิทยานิพนธ์ อาจารย์ ดร. สาวิตรี บุญพัชรนนท์ และ ศาสตราจารย์ Anthony Hayter กรรมการสอบวิทยานิพนธ์ ที่ได้กรุณาสละเวลาในการสอบวิทยานิพนธ์ และตรวจทานแก้ไขวิทยานิพนธ์ฉบับนี้ ตลอดทั้งให้คำแนะนำที่เป็นประโยชน์ที่ช่วยให้วิทยานิพนธ์ฉบับนี้สมบูรณ์มากขึ้น

ผู้วิจัยขอกราบขอบพระคุณคณาจารย์ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ทุกท่านที่ได้กรุณามอบองค์ความรู้ทางคณิตศาสตร์ และสถิติ ทำให้ผู้วิจัยสามารถนำมาประยุกต์ใช้ในวิทยานิพนธ์ฉบับนี้ ตลอดทั้งเจ้าหน้าที่ภาควิชาสถิติทุกท่านที่ได้อำนวยความสะดวกในด้านเอกสาร และการประสานงานต่าง ๆ ให้สำเร็จลุล่วงไปด้วยดี

สุดท้ายนี้ขอกราบขอบพระคุณบิดา มารดา ครอบครัว และมิตรสหาย ที่คอยสนับสนุนในด้านกำลังใจ กำลังใจ มาตลอดระยะเวลาในการทำวิจัย จนงานวิจัยเสร็จสมบูรณ์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญ	ช
สารบัญตาราง	ฌ
สารบัญรูปภาพ	ญ
1 บทนำ	1
1.1 ความเป็นมาและความสำคัญของปัญหา	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ขอบเขตของการวิจัย	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ	3
2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง	4
2.1 โครงข่ายประสาทแบบป้อนไปหน้า (Feedforward Neural Network; FNN)	4
2.2 วิธีถดถอยตามบริบท (Contextual regression method; CR)	7
2.3 วิธีถดถอยตามบริบทนัยทั่วไป (Generalized contextual regression method; GCR)	8
2.4 ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic tangent function; Tanh)	9
2.5 วิธีหาค่าที่ดีที่สุดที่สุดอคติม (ADAM-optimizer)	9
3 วิธีดำเนินงานวิจัย	11
3.1 การจำลองข้อมูล	11

	ช หน้า
3.2	เงื่อนไขที่ทำการศึกษา 18
3.3	แนวคิดในการออกแบบแต่ละตัวแบบ 18
3.4	ขั้นตอนในการดำเนินงานวิจัย 28
3.5	แผนผังแสดงขั้นตอนการทำงานในแต่ละกรณีทดลอง 29
4	ผลการวิจัย 30
4.1	กรณี 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น 30
4.2	กรณี 3 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น 32
4.3	กรณี 2 บริบท 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว 34
4.4	กรณี 3 บริบท 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว 38
4.5	ผลการวิเคราะห์จุดแบ่งระหว่างบริบทสำหรับตัวแบบ GCR 42
5	สรุปผลการวิจัยและข้อเสนอแนะ 44
5.1	สรุปผลการวิจัย 44
5.2	ข้อเสนอแนะ 46
ประวัติผู้เขียนวิทยานิพนธ์ 47	
รายการอ้างอิง 48	

สารบัญตาราง

ตารางที่	หน้า
4.1 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.1	31
4.2 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.2	33
4.3 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.3	35
4.4 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.3	37
4.5 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.4	39
4.6 ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.4	41

สารบัญรูปภาพ

รูปที่	หน้า
2.1 ตัวอย่างโครงสร้างของ FNN ที่มีชั้นที่ซ่อนอยู่ 2 ชั้น	4
2.2 แนวคิดในการสร้าง Contextual regression model อ้างอิงจาก [5] โดยมีการ หาค่าถ่วงน้ำหนักเชิงบริบท c_i สำหรับแต่ละตัวแปร x_i จากการวิเคราะห์ผ่านตัวแบบ ประเภทต่าง ๆ เช่นโครงข่ายประสาทเทียม และซัพพอร์ตเวกเตอร์แมชชีน	7
3.1 กราฟสีแดงแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 1 ($x_2 \geq 0$) และกราฟ สีน้ำเงินแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 2 ($x_2 < 0$)	12
3.2 กราฟสีแดงแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 1 ($x_2 \geq \frac{1}{3}$) , กราฟ สีน้ำเงินแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 2 ($\frac{1}{3} > x_2 \geq -\frac{1}{3}$) และ กราฟสีเขียวแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 3 ($x_2 < -\frac{1}{3}$)	14
3.3 โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 1	20
3.4 โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 2	21
3.5 โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 3.1	22
3.6 โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 3.2	23
3.7 โครงสร้าง neural network ของตัวแบบ FNN และตัวแบบ CR สำหรับกรณีศึกษาที่ 4.1	24
3.8 โครงสร้าง neural network ของตัวแบบ GCR สำหรับกรณีศึกษาที่ 4.1	25
3.9 โครงสร้าง neural network ของตัวแบบ FNN และตัวแบบ CR สำหรับกรณีศึกษาที่ 4.2	26
3.10 โครงสร้าง neural network ของตัวแบบ GCR สำหรับกรณีศึกษาที่ 4.2	27
4.1 กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	31
4.2 (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของ แต่ละตัวแบบ	32
4.3 กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	33
4.4 (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของ แต่ละตัวแบบ	34

4.5	กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	35
4.6	(บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ	36
4.7	กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	37
4.8	(บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ	38
4.9	กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	39
4.10	(บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ	40
4.11	กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ	41
4.12	(บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ [1.05ex](ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ	42
4.13	กราฟแสดงการกระจายตัวของค่า ω_6	43
5.1	แนวคิดในการรวบโหนดที่เกี่ยวข้องกับการวิเคราะห์ตามบริบทให้กลายเป็นโหนดเดียว . .	46

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน ด้วยความก้าวหน้าทางเทคโนโลยี ประกอบกับองค์ความรู้ทางการคำนวณที่เพิ่มขึ้นเรื่อย ๆ ทำให้เราสามารถวิเคราะห์ปัญหาที่เกี่ยวข้องกับข้อมูลขนาดใหญ่ได้ โดยปัญหาการเรียนรู้แบบมีผู้สอน (Supervised learning) ถือเป็นปัญหาหลัก ๆ ที่เจอได้ในหลากหลายอุตสาหกรรม ซึ่งปัญหาประเภทนี้ สามารถแก้ได้โดยการพยายามหาความสัมพันธ์บางอย่างระหว่างข้อมูลกับผลเฉลย ซึ่งความสัมพันธ์ดังกล่าวก็สามารถเป็นได้หลายรูปแบบ เช่น ความสัมพันธ์เชิงเส้น, ความสัมพันธ์ไม่เชิงเส้น, ความสัมพันธ์เชิงความน่าจะเป็น, ฯลฯ โดยในวิทยานิพนธ์ฉบับนี้จะกล่าวถึงข้อมูลที่มีความสัมพันธ์เชิงบริบท (Contextual relation)

การที่ข้อมูลมีความสัมพันธ์เชิงบริบท หมายถึงมีข้อมูลบางค่าที่เป็นตัวสร้างเงื่อนไขในการวิเคราะห์โมเดลความสัมพันธ์ระหว่างข้อมูลกับผลเฉลย ตัวอย่างเช่น ในสถานการณ์ปัจจุบัน ท่ามกลางปัญหาโรคโควิด 19 ผู้ป่วยจะต้องมีการวัดระดับออกซิเจนที่ปลายนิ้วมือ ซึ่งหลักการทำงานของเครื่องวัดนี้ จะใช้หลักการดูดกลืนแสง โดยมีแสงสีแดงส่องผ่านปลายนิ้วมือนี้ไปยังส่วนที่รับแสงด้านล่าง โดยเครื่องจะนำค่าปริมาณแสงที่วัดได้ก่อน และหลังผ่านนิ้วมือ ไปคำนวณค่าความอิ่มตัวของออกซิเจน และแสดงผ่านทางจอแสดงผล ซึ่งจากงานวิจัยของ Michael W. Sjoding [8] พบว่า ความแม่นยำในการวัดผลนั้นขึ้นอยู่กับความเข้มสีผิวของผู้ป่วย โดยผู้ป่วยที่มีสีผิวเข้มจะมีค่าออกซิเจนที่วัดได้สูงกว่าค่าจริง เมื่อวัดค่าจริงจากวิธี arterial blood gas จากข้อมูลนี้ เราจะกล่าวว่ามีสีผิวเป็นตัวแปรเชิงบริบท (Contextual features) เพราะเราสามารถเขียนสูตรการคำนวณหาความเข้มออกซิเจนได้เป็น

$$\text{ปริมาณความอิ่มตัวของ oxygen} = \begin{cases} model1(input) & \text{เมื่อความเข้มสีผิวมาก} \\ model2(input) & \text{เมื่อความเข้มสีผิวน้อย} \end{cases}$$

ซึ่งหากเราสามารถพัฒนาอุปกรณ์ให้คำนวณค่าความอิ่มตัวของ oxygen ได้ตามตัวแบบข้างต้น จะทำให้เราสามารถเพิ่มประสิทธิภาพของอุปกรณ์ได้มากขึ้นไปอีก โดยนอกเหนือจากตัวอย่างข้างต้นที่กล่าวมานั้น ยังมีอีกหลายปัญหาที่ประกอบด้วยข้อมูลที่มีความสัมพันธ์เชิงบริบท เช่น การวิเคราะห์คะแนนสอบของนักศึกษามหาวิทยาลัยของ Josée Dupuis [3] ที่มีการวิเคราะห์ตัวแปรเชิงบริบทเป็นการดูสื่อวีดิโอออนไลน์ของนักศึกษา และการสร้างตัวแบบอัตราการเติบโตของเด็กเล็กของ Laura M. Grajeda [4]

ที่มีการเคราะห์ตัวแปรเชิงบริบทเป็นอายุ และเพศของเด็ก

การแก้ปัญหาการเรียนรู้แบบมีผู้สอนนั้นมีได้หลายวิธีตั้งแต่ การถดถอยเชิงเส้น (LR; Linear regression), การถดถอยเชิงเส้นพหุคูณ (MLR; Multiple linear regression), Support vector machine (SVM) หรือโครงข่ายประสาทเทียม (ANN; Artificial Neural Networks) ซึ่งวิธี ANN นั้นเป็นที่นิยมในปัจจุบัน เนื่องจากสามารถแก้ไขปัญหามีความซับซ้อนได้ ซึ่งมีตัวอย่างการใช้งานคือ การแยกประเภทต้นไม้ว่าเป็นโรคใบไม่ร่วงหรือเปล่า ของ Noor Ezan Abdullah [1] หรือการทำนายเส้นผ่านศูนย์กลางต้นยูคาลิปตัสเพื่อหาปริมาตร ของ Fabrizzio Alphonsus [9] โดยลักษณะของปัญหาประเภทหนึ่งที่สามารถนำ ANN มาช่วยแก้ได้คือปัญหาประเภทการถดถอย ซึ่งเราสามารถใส่โครงสร้างพื้นฐานอย่าง Feedforward Neural Networks (FNN) ในการหาความสัมพันธ์ระหว่างข้อมูล และผลเฉลยได้ เช่นการพยากรณ์ความต้องการไฟฟ้า ของ M. Talaat [10] แต่ทั้งนี้ FNN ยังไม่สามารถแก้ไขปัญหามีข้อมูลเชิงบริบทได้อย่างมีประสิทธิภาพเท่าที่ควร

ในปี 2017 Chengyu Liu ได้เสนอวิธี การวิเคราะห์การถดถอยตามบริบท (Contextual regression method; CR) [5] ซึ่งสามารถแก้ปัญหาข้อมูลที่มีความสัมพันธ์เชิงบริบทกับผลเฉลยได้ โดยมีตัวอย่างการนำวิธีดังกล่าวไปใช้ในงานวิจัยของ Jonathan Parkinson [6]

ในวิทยานิพนธ์ฉบับนี้จะนำเสนอวิธีการแก้ปัญหาข้อมูลที่มีความสัมพันธ์เชิงบริบทกับผลเฉลย ในอีกรูปแบบโดยมีชื่อว่า การวิเคราะห์การถดถอยตามบริบทนัยทั่วไป (Generalized contextual regression method; GCR) ซึ่งมีวิธีการที่คล้ายคลึงกันกับงานวิจัยของ Chengyu Liu [5] ข้างต้น แต่มีการปรับโครงสร้างของตัวแบบใหม่เพื่อให้อยู่ในรูปทั่วไปมากขึ้น โดยจะมีการนำผลที่ได้จากตัวแบบที่ปรับโครงสร้างขึ้นมาี้ เทียบกับตัวแบบ FNN และ CR

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อประยุกต์ใช้ตัวแบบ GCR ในการแก้ไขปัญหาลinear regression ที่มีตัวแปรเชิงบริบท
2. เพื่อเปรียบเทียบผลลัพธ์ที่ได้จากการใช้ตัวแบบ FNN, ตัวแบบ CR และตัวแบบ GCR

1.3 ขอบเขตของการวิจัย

1. ศึกษาปัญหาประเภท linear regression เท่านั้น
2. ศึกษาปัญหา contextual features ที่มีความสัมพันธ์กับผลเฉลยในรูปแบบ linear และ non-linear
3. ปัญหาประกอบด้วย ตัวแปรอย่างน้อย 2 ตัวไม่เกิน 10 ตัว
4. ปัญหาประกอบด้วย ตัวแปรเชิงบริบทอย่างน้อย 1 ตัว แต่ไม่เกิน 3 ตัว
5. ปัญหาประกอบด้วยบริบทจำนวนไม่เกิน 3 บริบท
6. ข้อมูลที่เกิดจากการจำลองมีขนาดตัวอย่างเท่ากับ 1000 ตัวอย่าง ต่อ 1 บริบท
7. บริเวณของแต่ละบริบทที่ขึ้นกับฟังก์ชันเชิงบริบท มีลักษณะเป็นช่วง (Interval)

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อเป็นแนวทางในการประยุกต์ใช้โครงข่ายประสาทเทียมสำหรับการวิเคราะห์การถดถอยเชิงเส้นตามบริบทน้อยทั่วไป (GCR) ในการแก้ไขปัญหาค่าความสัมพันธ์เชิงบริบท
2. เพื่อได้มาซึ่งผลลัพธ์การเปรียบเทียบตัวแบบ FNN, CR และ GCR เมื่อนำมาประยุกต์ใช้กับการแก้ปัญหาค่าความสัมพันธ์เชิงบริบท

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 โครงข่ายประสาทแบบป้อนไปหน้า

(Feedforward Neural Network; FNN)

FNN เป็นโครงสร้างรูปแบบหนึ่งของ ANN มีลักษณะเป็นหลาย ๆ ชั้น (layer) โดยแต่ละชั้นจะประกอบด้วย หลาย ๆ โหนด (node) โดยทั่วไปแล้วโครงสร้างของ layer จะมีอยู่ 3 ส่วนหลัก ๆ คือ

1. ชั้นข้อมูลเข้า (input layer)

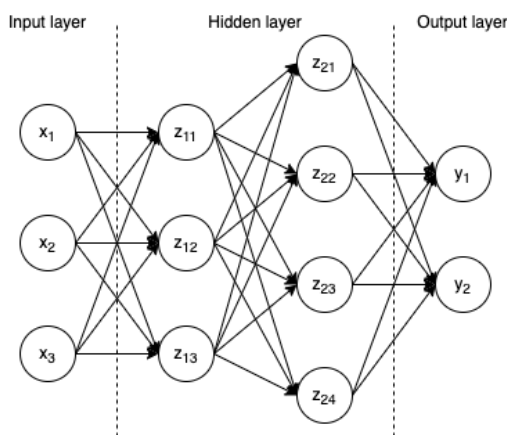
เป็นชั้นที่ประกอบด้วยโหนดที่เป็นข้อมูลเข้า ที่จะนำมาวิเคราะห์

2. ชั้นที่ซ่อนอยู่ (hidden layer)

เป็นชั้นที่ประกอบด้วยโหนดที่ทำหน้าที่เป็นตัวแปลงข้อมูลจากชั้นข้อมูลเข้ามาสสร้างค่าใหม่ (feature extraction) ก่อนที่จะส่งผ่านไปยังชั้นข้อมูลออก ซึ่งในนี้ อาจจะประกอบไปด้วยมากกว่า 1 ชั้นก็ได้

3. ชั้นข้อมูลออก (output layer)

เป็นชั้นที่ประกอบด้วยโหนดที่เป็นผลลัพธ์สุดท้ายจากการผ่านชั้นต่าง ๆ ก่อนหน้า



รูปที่ 2.1: ตัวอย่างโครงสร้างของ FNN ที่มีชั้นที่ซ่อนอยู่ 2 ชั้น

จากรูปที่ 2.1 เป็นตัวอย่างของ FNN ที่ประกอบด้วยชั้นข้อมูลเข้าที่มี 3 โหนด, ชั้นที่ซ่อนอยู่ 2 ชั้น ที่มี 3 และ 4 โหนดตามลำดับ และชั้นข้อมูลออกที่มี 2 โหนด โดยสำหรับกระบวนการทำงานของ FNN จะมี 2 ส่วนหลัก ๆ คือ

1. Feedforward

คือกระบวนการส่งผ่านข้อมูลไปด้านหน้า เมื่อรู้ค่าถ่วงน้ำหนัก (weight) ในแต่ละชั้น เช่นจากรูปที่ 2.1 จะได้ว่า

ชั้นที่ 1:

$$z_{1i} = f(\omega_{1i0} + \omega_{1i1} \cdot x_1 + \omega_{1i2} \cdot x_2 + \omega_{1i3} \cdot x_3)$$

สำหรับ $i = 1, 2, 3$

ชั้นที่ 2:

$$z_{2i} = f(\omega_{2i0} + \omega_{2i1} \cdot z_{11} + \omega_{2i2} \cdot z_{12} + \omega_{2i3} \cdot z_{13})$$

สำหรับ $i = 1, 2, 3, 4$

ชั้นข้อมูลออก:

$$y_i = f(\omega_{3i0} + \omega_{3i1} \cdot z_{21} + \omega_{3i2} \cdot z_{22} + \omega_{3i3} \cdot z_{23} + \omega_{3i4} \cdot z_{24})$$

สำหรับ $i = 1, 2$

เมื่อ f เป็นฟังก์ชันกระตุ้น (activation function) ซึ่งอาจจะเป็น sigmoid function หรือ tanh หรือ ReLU ขึ้นกับความเหมาะสมในการใช้งาน และ $w_{ijk} \in \mathbb{R}$ เป็นค่าถ่วงน้ำหนักที่เหมาะสม ซึ่งหากรู้ค่าถ่วงน้ำหนักในแต่ละชั้น จะทำให้เราสามารถประมาณค่าผลลัพธ์ได้จากกระบวนการ feedforward นี้

2. Backpropagation

เนื่องจากในทางปฏิบัติแล้ว เราจะไม่สามารถหาค่าถ่วงน้ำหนักได้โดยตรง ดังนั้นกระบวนการ backpropagation จะเป็นวิธีที่ช่วยเราในการหาค่าถ่วงน้ำหนักที่เหมาะสม โดยใช้การเคลื่อนลงตามความชัน (gradient descent) เข้ามาช่วย โดยมีวิธีการดังนี้

- กำหนดรูปทั่วไปของ \hat{y} ให้ขึ้นกับ ข้อมูลเข้า x และค่าถ่วงน้ำหนักในแต่ละชั้น ω หรือก็คือเขียนให้อยู่ในรูป $\hat{y} = h(x, \omega)$

- b) กำหนดฟังก์ชันการสูญเสีย (loss function) $L_{x,y}(\omega)$ เช่นหากกำหนดให้เป็นค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean squared error) จะได้เป็น

$$L_{x,y}(\omega) = (\hat{y} - y)^2 = (h(x, \omega) - y)^2$$

- c) หา ω ที่ทำให้ค่า $L_{x,y}(\omega)$ มีค่าน้อยสุด โดยมีวิธีการหาหลากหลายวิธี เช่น Batch gradient descent (BGD) ตามอัลกอริทึม 1

Algorithm 1 Batch Gradient Descent Algorithm

```

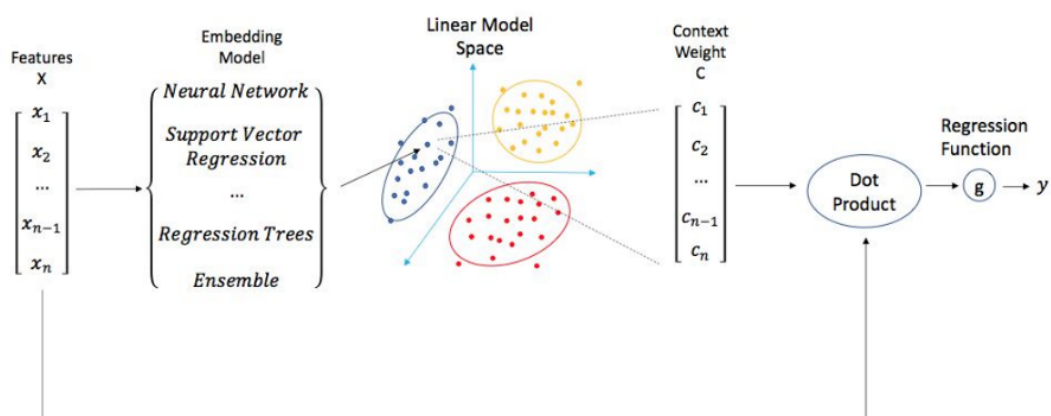
1: Initiate  $\omega^0, \epsilon$ 
2:  $t \leftarrow 1$ 
3: while true do
4:   set learning rate  $\eta_t$ 
5:    $\omega^t \leftarrow \omega^{t-1} - \eta_t \nabla L(\omega^{t-1})$ 
6:   if  $\|\omega^t - \omega^{t-1}\| < \epsilon$  then
7:      $\omega^* \leftarrow \omega^t$ ;  $\omega^*$  is the optimal weight
8:     break
9:   end if
10:   $t \leftarrow t + 1$ 
11: end while

```

2.2 วิธีถดถอยตามบริบท

(Contextual regression method; CR)

Chengyu Liu. และ Wei Wang. ได้มีการเสนอตัวแบบ contextual regression เพื่อช่วยแก้ปัญหาเรื่องการตีความได้ของตัวแบบ และมีความแม่นยำไปพร้อม ๆ กัน โดยมีแนวคิดดังรูปที่ 2.2 ซึ่งจะมีการสร้างเวกเตอร์ $c(x)$ ที่เป็นค่าถ่วงน้ำหนักเชิงบริบท แล้วนำมาเป็นส่วนช่วยในการคำนวณ



รูปที่ 2.2: แนวคิดในการสร้าง Contextual regression model อ้างอิงจาก [5] โดยมีการหาค่าถ่วงน้ำหนักเชิงบริบท c_i สำหรับแต่ละตัวแปร x_i จากการวิเคราะห์ผ่านตัวแบบประเภทต่าง ๆ เช่น โครงข่ายประสาทเทียม และซัพพอร์ตเวกเตอร์แมชชีน

จากรูปที่ 2.2 สามารถเขียนออกมาเป็นตัวแบบในรูปทั่วไปได้เป็น

$$y = g(c(x)^t \cdot x + b) = g\left(\sum_{i=1}^n c_i(x_1, x_2, x_3, \dots, x_n) \cdot x_i + b\right) \quad (2.1)$$

เมื่อ g เป็นฟังก์ชันการถดถอย เช่น ฟังก์ชันการระบุตัวตน (Identity function) สำหรับการถดถอยเชิงเส้น หรือฟังก์ชันโลจิสติก (Logistic function) สำหรับการถดถอยโลจิสติก โดยในตัวแบบนี้มีการออกแบบการหาค่าถ่วงน้ำหนักตามบริบท (Contextual weight) c_1, c_2, \dots, c_n ด้วยวิธี Bidirectional LSTM โดยเทียบผลลัพธ์กับวิธีต่าง ๆ เช่น Linear Regression, Lasso Regression, LSTM (log input) ซึ่งจากผลการทดลองของ Chengyu Liu. และ Wei Wang. พบว่าวิธี Contextual regression (log input) ให้ประสิทธิภาพที่ดีที่สุด

2.3 วิธีถดถอยตามบริบททั่วไ

(Generalized contextual regression method; GCR)

ในวิทยานิพนธ์ฉบับนี้จะมีการตีความหมาย contextual regression ในอีกลักษณะหนึ่ง ซึ่งจะเขียนตัวแบบออกมาในหน้าตาที่แตกต่างกัน แต่สุดท้ายสามารถจัดรูปออกมาให้สมมูลกันกับสมการ 2.1 ได้ในบางกรณี โดยแนวทางการสร้างเริ่มจากกำหนดให้ความสัมพันธ์ทั้งหมด p บริบท โดยจะกล่าวว่าเป็นบริบทที่ 1, บริบทที่ 2, ..., บริบทที่ p และกำหนดให้แต่ละตัวอย่างประกอบด้วย $y \in \mathbb{R}$ เป็นตัวแปรตาม, $x_{11}, x_{12}, \dots, x_{1m} \in \mathbb{R}$ เป็นตัวแปรอิสระทั่วไป และ $x_{21}, x_{22}, \dots, x_{2n} \in \mathbb{R}$ เป็นตัวแปรอิสระเชิงบริบท

การที่ $x_{21}, x_{22}, \dots, x_{2n}$ เป็นตัวแปรเชิงบริบท สมมติให้สอดคล้องกับบริบทที่ i หมายความว่า มีฟังก์ชัน g ที่ส่งค่า $x_{21}, x_{22}, \dots, x_{2n}$ ไปยังบริบทที่ i หรือเขียนได้เป็น

$$\exists g : \mathbb{R}^n \rightarrow \mathbb{R}, g(x_{21}, x_{22}, \dots, x_{2n}) \in \mathcal{U}_i$$

เมื่อ g คือฟังก์ชันเชิงบริบท (Contextual function) และ \mathcal{U}_i คือเซตของภาพฉายของ g ที่เกิดจาก $x_{21}, x_{22}, \dots, x_{2n}$ ที่สอดคล้องกับบริบทที่ i หรือจะเรียกว่าเป็นบริเวณของบริบทที่ i (i^{th} Contextual space) เมื่อ $\bigcup_{i=1}^p \mathcal{U}_i = \mathbb{R}$ และ $\bigcap_{i=1}^p \mathcal{U}_i = \emptyset$

สำหรับแต่ละตัวอย่าง จะมีตัวแบบแตกต่างกันตามแต่ละบริบท สมมติให้ $x_{11}, x_{12}, \dots, x_{1m}$ สอดคล้องกับบริบทที่ i จะมีตัวแบบเป็น $y = f_i(x_{11}, x_{12}, \dots, x_{1m}, x_{21}, x_{22}, \dots, x_{2n})$ ซึ่งเราสามารถเขียนตัวแบบในรูปทั่วไปออกมาได้เป็น

$$y = \sum_{i=1}^p I(g(x_{21}, x_{22}, \dots, x_{2n}) \in \mathcal{U}_i) \cdot f_i(x_{11}, x_{12}, \dots, x_{1m}, x_{21}, x_{22}, \dots, x_{2n})$$

เมื่อ $I(\cdot)$ เป็นฟังก์ชันบ่งชี้ (Indicator function) กล่าวคือ

$$I(g(x_{21}, x_{22}, \dots, x_{2n}) \in \mathcal{U}_i) = \begin{cases} 1 & \text{เมื่อ } g(x_{21}, x_{22}, \dots, x_{2n}) \in \mathcal{U}_i \\ 0 & \text{เมื่อ } g(x_{21}, x_{22}, \dots, x_{2n}) \notin \mathcal{U}_i \end{cases}$$

ซึ่งหากให้ $x = (x_{11}, x_{12}, \dots, x_{1m}, x_{21}, x_{22}, \dots, x_{2n})^t$ และจัดรูปให้

$g(x_{21}, x_{22}, \dots, x_{2n}) = g(x_{11}, x_{12}, \dots, x_{1m}, x_{21}, x_{22}, \dots, x_{2n}) = g(x)$ จะได้ว่า

$$y = \sum_{i=1}^p I(g(x) \in \mathcal{U}_i) \cdot f_i(x) \quad (2.2)$$

โดยในที่นี้เราจะกล่าวว่าตัวแบบเป็น

1. บริบทเชิงเส้น (linear contextual) เมื่อ g เป็นฟังก์ชันเชิงเส้น หรือก็คือมี $\omega \in \mathbb{R}^{m+n}$ และ $b \in \mathbb{R}$ ที่ทำให้ $g(x) = \omega^t x + b$
2. บริบทไม่เชิงเส้น (nonlinear contextual) เมื่อ g เป็นฟังก์ชันไม่เชิงเส้น ซึ่งมีหลายหลายรูปแบบ เช่น $g(x) = x^t W x$ เมื่อ $W \in \mathbb{R}^{(m+n) \times (m+n)}$

จากสมการ 2.2 พบว่าถ้าให้ $f_i(x) = a_i^t x + b_i$ เมื่อ $a_i \in \mathbb{R}^{m+n}, b_i \in \mathbb{R}$ จะได้ว่า

$$\begin{aligned} y &= \sum_{i=1}^p I(g(x) \in \mathcal{U}_i) \cdot (a_i^t x + b_i) \\ &= \underbrace{\left(\sum_{i=1}^p I(g(x) \in \mathcal{U}_i) a_i^t \right)}_{c(x)^t} \cdot x + \underbrace{\sum_{i=1}^p I(g(x) \in \mathcal{U}_i) b_i}_b \end{aligned}$$

ซึ่งสอดคล้องกับรูปแบบที่ Chengyu Liu และ Wei Wang เสนอไว้ แต่ตัวแบบ GCR จะมีลักษณะทั่วไปกว่า เพราะไม่ได้เจาะจงว่าแต่ละ f_i จากสมการ 2.2 จะต้องเป็นฟังก์ชันเชิงเส้นเท่านั้น สามารถกำหนดให้เป็นฟังก์ชันไม่เชิงเส้นได้ และมากกว่านั้นตัวแบบ GCR มีความสามารถในการอธิบายข้อมูลได้ดีกว่าตัวแบบ CR ผ่านการวิเคราะห์บริเวณของแต่ละบริบท ซึ่งทั้งสองส่วนนี้ถือเป็นข้อแตกต่างที่เพิ่มขึ้นมาในตัวแบบ GCR

2.4 ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic tangent function; Tanh)

โดยทั่วไปแล้ว การส่งค่าผ่านในแต่ละชั้นของโครงข่ายประสาทเทียม จะมีการนำข้อมูลไปผ่านฟังก์ชันกระตุ้น (activation function) ที่เป็นฟังก์ชันไม่เชิงเส้นก่อนจะส่งไปยังชั้นถัดไป เพื่อเพิ่มความซับซ้อนให้กับตัวแบบ โดยในงานวิจัยฉบับนี้เลือกใช้ฟังก์ชันไฮเพอร์โบลิกแทนเจนต์ (Hyperbolic tangent function; tanh) ซึ่งเป็นหนึ่งในฟังก์ชันกระตุ้นหลาย ๆ ฟังก์ชันอ้างอิงจาก [7] โดยมีลักษณะดังสมการ

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

เมื่อ $x \in \mathbb{R}$ โดยค่าที่ได้จากฟังก์ชันไฮเพอร์โบลิกแทนเจนต์จะอยู่ในช่วง -1 ถึง 1

2.5 วิธีหาค่าที่ดีที่สุดอดัม (ADAM-optimizer)

ในการทำ backpropagation ของตัวแบบโครงข่ายประสาทเทียม มีวัตถุประสงค์เพื่อหาค่าถ่วงน้ำหนักที่ทำให้ฟังก์ชันการสูญเสียให้ค่าน้อยที่สุด โดยมีวิธีหาค่าที่ดีที่สุด (optimizer) หลายตัวที่สามารถนำมาใช้ได้ โดยในงานวิจัยฉบับนี้จะเลือกใช้วิธีหาค่าที่ดีที่สุดอดัม [2] เพราะเป็นวิธีหาค่าที่ดีที่สุดที่มีการนำข้อดีจากวิธีหาค่าที่ดีที่สุดอื่น ๆ มารวมกัน โดยมี algorithm ตามอัลกอริทึม 2 ด้านล่าง

Algorithm 2 Adaptive Moment Estimation (ADAM)

- 1: Initiate $t \leftarrow 0$, number of iterations as T , learning rate γ and $\epsilon > 0$
 - 2: Define loss function $L(\omega)$ for each ω
 - 3: ω_0 be a random unit vector
 - 4: $m_0 \leftarrow 0$ (first moment)
 - 5: $v_0 \leftarrow 0$ (second moment)
 - 6: **while** $t < T$ **do**
 - 7: $t \leftarrow t + 1$
 - 8: $g_t \leftarrow \nabla_{\omega} L(\omega_{t-1})$
 - 9: $m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$
 - 10: $v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$
 - 11: $\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$
 - 12: $\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$
 - 13: $w_t \leftarrow w_{t-1} - \gamma \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$
 - 14: **end while**
 - 15: **return** w_T
-

บทที่ 3

วิธีดำเนินงานวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบโครงสร้างตัวแบบ GCR ที่ผู้วิจัยพัฒนาขึ้นมา ว่ามีประสิทธิภาพเทียบดีกว่าตัวแบบดั้งเดิมอย่าง FNN และ CR หรือไม่ อย่างไร โดยผ่านการจำลองข้อมูลที่มีความสัมพันธ์เชิงบริบทในรูปแบบเชิงเส้น และไม่เชิงเส้น

3.1 การจำลองข้อมูล

ทางผู้วิจัยได้ทำการวิเคราะห์ปัญหา จากการจำลองข้อมูลที่มีความสัมพันธ์เชิงเส้นตามบริบททั้งหมด 4 รูปแบบ และมีวิธีการจำลองข้อมูลในแต่ละกรณีดังนี้

1. กรณี 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)
 - 1.1. กำหนดขนาดตัวอย่างเท่ากับ 2000 ตัวอย่าง โดยมีบริบทละ 1000 ตัวอย่าง
 - 1.2. กำหนดให้ตัวแปรอิสระ 2 ตัวคือ x_1, x_2 ที่มี x_2 เป็นตัวแปรเชิงบริบท โดย $x_1, x_2 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$
 - 1.3. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทเชิงเส้น (Linear contextual linear models)

$$y = \begin{cases} a_1 x_1 + b_1 + \epsilon_1 & \text{เมื่อ } x_2 \geq c \\ a_2 x_1 + b_2 + \epsilon_2 & \text{เมื่อ } x_2 < c \end{cases} \quad (3.1)$$

เมื่อ

y เป็นตัวแปรตาม

x_1 เป็นตัวแปรอิสระ

x_2 เป็นตัวแปรอิสระเชิงบริบท

a_1, b_1, a_2, b_2 เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1 และบริบทที่ 2 ตามลำดับ โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด ϵ_1, ϵ_2 เป็นตัวแปรรบกวน โดย $\epsilon_1, \epsilon_2 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

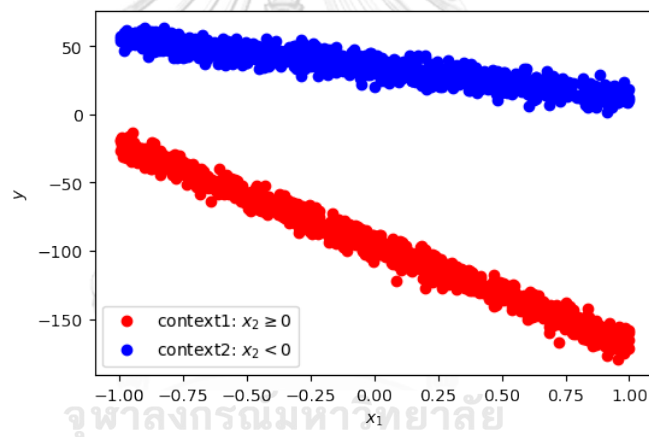
c คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1 และบริบทที่ 2 หรือก็คือ $U_1 = [c, \infty)$, $U_2 = (-\infty, c)$ โดยในที่นี้ กำหนดให้ $c = 0$

- 1.4. ในการทดลองได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบข้างต้นออกมาดังนี้ $a_1 = -73, a_2 = -22, b_1 = -96, b_2 = 35$ นั่นคือ

$$y = \begin{cases} -73x_1 - 96 + \epsilon_1 & \text{เมื่อ } x_2 \geq 0 \\ -22x_1 + 35 + \epsilon_2 & \text{เมื่อ } x_2 < 0 \end{cases} \quad (3.2)$$

เมื่อ $\epsilon_1, \epsilon_2 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

- 1.5. ทำการสุ่มค่า $x_1, x_2 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$ และ มา $\epsilon_1, \epsilon_2 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ 2000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1 และ บริบทที่ 2 บริบทละ 1000 ตัวอย่าง และ คำนวณค่า y ตามสมการ 3.2 โดยสามารถแสดงความสัมพันธ์ได้ดังรูปที่ 3.1



รูปที่ 3.1: กราฟสีแดงแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 1 ($x_2 \geq 0$) และกราฟสีน้ำเงินแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 2 ($x_2 < 0$)

2. กรณี 3 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)

2.1. กำหนดขนาดตัวอย่างเท่ากับ 3000 ตัวอย่าง (บริบทละ 1000 ตัวอย่าง)

2.2. กำหนดให้ตัวแปรอิสระ 2 ตัวคือ x_1, x_2 ที่มี x_2 เป็นตัวแปรเชิงบริบท โดย $x_1, x_2 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$

2.3. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทเชิงเส้น (Linear contextual linear models)

$$y = \begin{cases} a_1x_1 + b_1 + \epsilon_1 & \text{เมื่อ } x_2 \geq c_1 \\ a_2x_1 + b_2 + \epsilon_2 & \text{เมื่อ } c_1 > x_2 \geq c_2 \\ a_3x_1 + b_3 + \epsilon_3 & \text{เมื่อ } x_2 < c_2 \end{cases} \quad (3.3)$$

เมื่อ

y เป็นตัวแปรตาม

x_1 เป็นตัวแปรอิสระ

x_2 เป็นตัวแปรอิสระเชิงบริบท

$a_1, b_1, a_2, b_2, a_3, b_3$ เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 ตามลำดับ โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด

$\epsilon_1, \epsilon_2, \epsilon_3$ เป็นตัวแปรรบกวนโดย $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

c_1, c_2 คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 ตามลำดับ หรือก็คือ $\mathcal{U}_1 = [c_1, \infty), \mathcal{U}_2 = [c_2, c_1), \mathcal{U}_3 = (-\infty, c_2)$ โดยในที่นี้ให้ $c_1 = \frac{1}{3}$ และ $c_2 = -\frac{1}{3}$

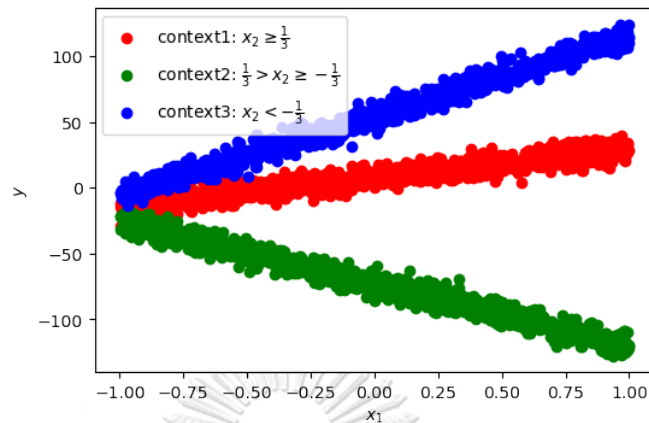
2.4. ในการทดลองได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบข้างต้นออกมาดังนี้ $a_1 = 22, a_2 = -47, a_3 = 60, b_1 = 8, b_2 = -73, b_3 = 55, c_1 = \frac{1}{3}$ นั่นคือ

$$y = \begin{cases} 22x_1 + 8 + \epsilon_1 & \text{เมื่อ } x_2 \geq \frac{1}{3} \\ -47x_1 - 73 + \epsilon_2 & \text{เมื่อ } \frac{1}{3} > x_2 \geq -\frac{1}{3} \\ 60x_1 + 55 + \epsilon_3 & \text{เมื่อ } x_2 < -\frac{1}{3} \end{cases} \quad (3.4)$$

เมื่อ $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

2.5. ทำการสุ่มค่า $x_1, x_2 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$ และ $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ มา 3000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 บริบทละ 1000

ตัวอย่าง และค่านวนค่า y ตามสมการ 3.4 จะทำให้ได้ชุดข้อมูล y, x_1, x_2 3000 ชุด โดยสามารถแสดงความสัมพันธ์ได้ดังรูปที่ 3.2



รูปที่ 3.2: กราฟสีแดงแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 1 ($x_2 \geq \frac{1}{3}$), กราฟสีน้ำเงินแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 2 ($\frac{1}{3} > x_2 \geq -\frac{1}{3}$) และกราฟสีเขียวแสดงความสัมพันธ์ระหว่าง y กับ x_1 ในบริบทที่ 3 ($x_2 < -\frac{1}{3}$)

3. กรณี 2 บริบท 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น หรือไม่เชิงเส้น)

3.1. กำหนดขนาดตัวอย่างเท่ากับ 2000 ตัวอย่าง (บริบทละ 1000 ตัวอย่าง)

3.2. กำหนดให้ $x = (x_1, x_2, x_3, x_4, x_5)^t$ โดยที่ x_4, x_5 เป็นตัวแปรเชิงบริบท ทั้งนี้ให้ $x' = (x_4, x_5)^t$ โดย $x_1, x_2, x_3, x_4, x_5 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$

3.3. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทเชิงเส้น (Linear contextual linear models)

$$y = \begin{cases} a_1^t x + b_1 + \epsilon_1 & \text{เมื่อ } \omega^t x' \geq c \\ a_2^t x + b_2 + \epsilon_2 & \text{เมื่อ } \omega^t x' < c \end{cases} \quad (3.5)$$

เมื่อ

y เป็นตัวแปรตาม

x_1, x_2, x_3 เป็นตัวแปรอิสระ

x_4, x_5 เป็นตัวแปรอิสระเชิงบริบท

a_1, b_1, a_2, b_2 เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1 และบริบทที่ 2 ตามลำดับ

โดยแต่ละพารามิเตอร์ใน a_1, b_1, a_2, b_2 โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด

ϵ_1, ϵ_2 เป็นตัวแปรปรบกวนโดย $\epsilon_1, \epsilon_2 \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

ω คือเวกเตอร์ขนาด 2×1 ที่ส่ง x' ไปยังแต่ละบริบท โดยแต่ละพารามิเตอร์ใน ω จะสุ่มค่าจาก $[-1, 1]$

c คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1 และบริบทที่ 2 ตามลำดับ โดยในที่นี้ให้ $c = 0$

3.4. สำหรับการทดลอง 3.3. ได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบออกมาดังนี้ $a_1 = [20, -83, 86, 38, -70]^t, a_2 = [-32, -60, 5, 80, -83]^t, b_1 = 38, b_2 = 46$ และ $\omega = [-0.25, 0.12]^t$

3.5. ทำการสุ่มค่า $x_1, x_2, x_3, x_4, x_5 \stackrel{i.i.d.}{\sim} \text{Cont.Unif}(-1, 1)$ และ $\epsilon_1, \epsilon_2 \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ มา 2000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1 และ บริบทที่ 2 บริบทละ 1000 ตัวอย่าง และคำนวณค่า y ตามสมการ 3.5 จะทำให้ได้ชุดข้อมูล $y, x_1, x_2, x_3, x_4, x_5$ 2000 ชุด

3.6. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทไม่เชิงเส้น (Nonlinear contextual linear models)

$$y = \begin{cases} a_1^t x + b_1 + \epsilon_1 & \text{เมื่อ } x'^t W x' \geq c \\ a_2^t x + b_2 + \epsilon_2 & \text{เมื่อ } x'^t W x' < c \end{cases} \quad (3.6)$$

เมื่อ

y เป็นตัวแปรตาม

x_1, x_2, x_3 เป็นตัวแปรอิสระ

x_4, x_5 เป็นตัวแปรอิสระเชิงบริบท

a_1, b_1, a_2, b_2 เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1 และบริบทที่ 2 ตามลำดับ โดยแต่ละพารามิเตอร์ใน a_1, b_1, a_2, b_2 โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด

ϵ_1, ϵ_2 เป็นตัวแปรปรบกวนโดย $\epsilon_1, \epsilon_2 \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

W คือเมทริกซ์ขนาด 2×2 ที่ส่ง x' ไปยังแต่ละบริบท โดยแต่ละ พารามิเตอร์ใน W จะสุ่มค่าจาก $[-1, 1]$ โดยมีเงื่อนไขว่าต้องเป็นเมทริกซ์สมมาตรที่เป็นบวกแน่นอน (Symmetric positive-definite matrix)

c คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1 และบริบทที่ 2 ตามลำดับ

3.7. สำหรับการทดลอง 3.6. ได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบออกมาดังนี้ $a_1 = [80, 63, 48, 97, 53]^t, a_2 = [74, -58, 53, -19, -8]^t, b_1 = -33, b_2 = -19, c =$

$$0.0904 \text{ และ } W = \begin{bmatrix} 0.6088 & -0.2001 \\ -0.2001 & 0.2421 \end{bmatrix}.$$

3.8. ทำการสุ่มค่า $x_1, x_2, x_3, x_4, x_5 \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$ และ $\epsilon_1, \epsilon_2 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ มา 2000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1 และ บริบทที่ 2 บริบทละ 1000 ตัวอย่าง และคำนวณค่า y ตามสมการ 3.6 จะทำให้ได้ชุดข้อมูล $y, x_1, x_2, x_3, x_4, x_5$ 2000 ชุด

4. กรณี 3 บริบท 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น หรือไม่เชิงเส้น)

4.1. กำหนดขนาดตัวอย่างเท่ากับ 3000 ตัวอย่าง (บริบทละ 1000 ตัวอย่าง)

4.2. กำหนดให้ $x = (x_1, x_2, x_3, \dots, x_{10})^t$ โดยที่ x_8, x_9, x_{10} เป็นตัวแปรเชิงบริบท ทั้งนี้ให้ $x' = (x_8, x_9, x_{10})^t$ โดย $x_1, x_2, x_3, \dots, x_{10} \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$

4.3. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทเชิงเส้น (Linear contextual linear models)

$$y = \begin{cases} a_1^t x + b_1 + \epsilon_1 & \text{เมื่อ } \omega^t x' \geq c_1 \\ a_2^t x + b_2 + \epsilon_2 & \text{เมื่อ } c_1 > \omega^t x' \geq c_2 \\ a_3^t x + b_3 + \epsilon_3 & \text{เมื่อ } \omega^t x' < c_2 \end{cases} \quad (3.7)$$

เมื่อ

y เป็นตัวแปรตาม

$x_1, x_2, x_3, \dots, x_7$ เป็นตัวแปรอิสระ

x_8, x_9, x_{10} เป็นตัวแปรอิสระเชิงบริบท

$a_1, b_1, a_2, b_2, a_3, b_3$ เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 ตามลำดับ โดยแต่ละพารามิเตอร์ใน $a_1, b_1, a_2, b_2, a_3, b_3$ โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด

$\epsilon_1, \epsilon_2, \epsilon_3$ เป็นตัวแปรรบกวนโดย $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

ω คือเวกเตอร์ขนาด 3×1 ที่ส่ง x' ไปยังแต่ละบริบท โดยแต่ละพารามิเตอร์ใน ω จะสุ่มค่าจาก $[-1, 1]$

c_1, c_2 คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1 ,บริบทที่ 2 และบริบทที่ 3 ตามลำดับ ในที่นี้ให้ $c_1 = 1$ และ $c_2 = -1$

4.4. สำหรับการทดลอง 4.3. ได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบออกมาดังนี้

$$a_1 = [-54, 1, 42, 27, 95, -15, 64, 96, 11, -12]^t,$$

$$a_2 = [-13, -38, 5, 24, 14, -6, 65, 87, -15, 88]^t,$$

$$a_3 = [-99, -49, -97, 68, -73, 22, -34, 89, -64, 68]^t,$$

$$b_1 = 12, b_2 = 35, b_3 = 37 \text{ และ } \omega = [-0.4583, -0.3444, 0.8089]^t$$

4.5. ทำการสุ่มค่า $x_1, x_2, x_3, \dots, x_{10} \stackrel{i.i.d.}{\sim} \text{Cont.Unif}(-1, 1)$ และ $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ มา 3000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 บริบทละ 1000 ตัวอย่าง และคำนวณค่า y ตามสมการ 3.7 จะทำให้ได้ชุดข้อมูล $y, x_1, x_2, x_3, \dots, x_{10}$ 3000 ชุด

4.6. ทำการจำลองข้อมูลที่มีความสัมพันธ์แบบบริบทไม่เชิงเส้น (Nonlinear contextual linear models)

$$y = \begin{cases} a_1^t x + b_1 + \epsilon_1 & \text{เมื่อ } x^t W x' \geq c_1 \\ a_2^t x + b_2 + \epsilon_2 & \text{เมื่อ } c_1 > x^t W x' \geq c_2 \\ a_3^t x + b_3 + \epsilon_3 & \text{เมื่อ } x^t W x' < c_2 \end{cases} \quad (3.8)$$

เมื่อ

y เป็นตัวแปรตาม

$x_1, x_2, x_3, \dots, x_7$ เป็นตัวแปรอิสระ

x_8, x_9, x_{10} เป็นตัวแปรอิสระเชิงบริบท

$a_1, b_1, a_2, b_2, a_3, b_3$ เป็นค่าสัมประสิทธิ์การถดถอยของบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 ตามลำดับ โดยแต่ละพารามิเตอร์ใน $a_1, b_1, a_2, b_2, a_3, b_3$ โดยสุ่มค่าแบบหยิบซ้ำได้จาก $\{-100, -99, \dots, 99, 100\}$ มา 1 ชุด

$\epsilon_1, \epsilon_2, \epsilon_3$ เป็นตัวแปรรบกวนโดย $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$

W คือเมทริกซ์ขนาด 3×3 ที่ส่ง x' ไปยังแต่ละบริบท โดยแต่ละพารามิเตอร์ใน W จะสุ่มค่าจาก $[-1, 1]$ โดยมีเงื่อนไขว่าต้องเป็นเมทริกซ์สมมาตรที่เป็นบวกแน่นอน (Symmetric positive-definite matrix)

c_1, c_2 คือจุดแบ่งระหว่าง บริเวณของบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 ตามลำดับ

4.7. สำหรับการทดลอง 4.6. ได้มีการสุ่มค่าพารามิเตอร์ของตัวแบบออกมาดังนี้

$$a_1 = [-69, 59, 23, 63, 25, -31, -70, 12, 56, -83]^t,$$

$$a_2 = [-73, 87, 70, 51, 92, -62, 28, 69, -62, -19]^t,$$

$$a_3 = [-63, -73, -91, -45, 8, 54, 35, -22, -55, 35]^t,$$

$$b_1 = -22, b_2 = -94, b_3 = 59, c_1 = 0.6427, c_2 = 0.3072 \text{ และ}$$

$$W = \begin{bmatrix} 0.2177 & -0.3952 & -0.2707 \\ -0.3952 & 0.9063 & 0.2663 \\ -0.2707 & 0.2663 & 0.6069 \end{bmatrix}.$$

- 4.8. ทำการสุ่มค่า $x_1, x_2, x_3, \dots, x_{10} \stackrel{i.i.d}{\sim} \text{Cont.Unif}(-1, 1)$ และ $\epsilon_1, \epsilon_2, \epsilon_3 \stackrel{i.i.d}{\sim} \mathcal{N}(\mu = 0, \sigma^2 = 25)$ มา 3000 ตัวอย่าง โดยสุ่มให้อยู่ในบริบทที่ 1, บริบทที่ 2 และบริบทที่ 3 บริบทละ 1000 ตัวอย่าง และคำนวณค่า y ตามสมการ 3.8 จะทำให้ได้ชุดข้อมูล $y, x_1, x_2, x_3, \dots, x_{10}$ 3000 ชุด

โดยจากการจำลองข้างต้นสังเกตว่ามีเพียงกรณี 1 และ 2 เท่านั้น ที่ตัวแปรเชิงบริบทไม่เป็นส่วนร่วมในความสัมพันธ์เชิงเส้น ทั้งนี้เพื่อต้องการสังเกตลักษณะที่เกิดขึ้นจากข้อมูลที่มีรูปแบบง่ายที่สุด

3.2 เงื่อนไขที่ทำการศึกษา

สำหรับการทำขั้นตอน backpropagation เพื่อให้สามารถเปรียบเทียบผลที่เกิดขึ้นในแต่ละตัวแบบได้โดยไม่เกิดความลำเอียง ดังนั้นแต่ละตัวแบบจะมีการกำหนดค่าพารามิเตอร์ที่สำคัญให้เป็นค่าเดียวกันดังนี้

1. อัตราการเรียนรู้ (learning rate) = 0.01
2. วิธีหาค่าที่ดีที่สุด (optimizer) เลือกเป็น Adam optimizer
3. จำนวนรอบในการทำ backpropagation ต่อ 1 ตัวแบบ (epoch) คือ 20000 รอบ
4. ฟังก์ชันกระตุ้นที่ใช้ในแต่ละตัวแบบคือ ฟังก์ชันไฮเพอร์โบลิกแทน (Tanh function)
5. ฟังก์ชันสูญเสียที่ใช้ในกระบวนการ backpropagation คือ ฟังก์ชันค่าเฉลี่ยความผิดพลาดกำลังสอง

3.3 แนวคิดในการออกแบบแต่ละตัวแบบ

สำหรับแต่ละกรณีศึกษา มีวิธีการสร้างตัวแบบที่แตกต่างกันตามความซับซ้อนของข้อมูลในแต่ละกรณี โดยมีแนวคิดในการสร้างตัวแบบในแต่ละกรณีดังนี้

กรณีที่ 1: 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)

1. Feedforward Neural Networks (FNN)

สำหรับกรณีนี้ เราต้องการสร้าง Neural network ที่มีความสามารถในการจำแนก 2 บริบท และสามารถสร้างความสัมพันธ์เชิงเส้น 2 รูปแบบ (บริบทละ 1 รูปแบบ) จึงมีการกำหนดให้ชั้นที่ซ่อนอยู่ที่ 1 ประกอบไปด้วย 3 โหนด โดยโหนดที่ 1 มีไว้เพื่อจำแนกบริบทจากข้อมูล, โหนดที่ 2 มีไว้สำหรับการทำนายผลเฉลยของบริบทที่ 1, โหนดที่ 3 มีไว้สำหรับการทำนายผลเฉลยของบริบทที่ 2 และชั้นที่ซ่อนอยู่ที่ 2 ประกอบไปด้วย 2 โหนด เพื่อเพิ่มความซับซ้อนของตัวแบบก่อนที่จะเข้าสู่ชั้นข้อมูลออก โดยจะมีโครงสร้างดังรูปที่ 3.3(a)

2. Contextual regression method (CR)

สำหรับกรณีนี้ เราจะนำเอาตัวแบบของ Chengyu Liu., Wei Wang. [5] ซึ่งได้แสดงไว้ดังสมการที่ 2.1 โดยในที่นี้เราจะแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบท(c_i) ในสมการดังกล่าวด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 2 โหนด เพื่อให้มีความซับซ้อนใกล้เคียงกันกับตัวแบบ FNN โดยจะมีโครงสร้างดังรูปที่ 3.3(b)

3. Generalized contextual regression method (GCR)

สืบเนื่องมาจากสมการ 2.2 หากเราจัดรูปใหม่ให้สอดคล้องกับสมการนี้จะได้ว่า

$$y = I(x_2 \geq 0) \cdot (-73x_1 - 96) + I(x_2 < 0) \cdot (-22x_1 + 35) \quad (3.9)$$

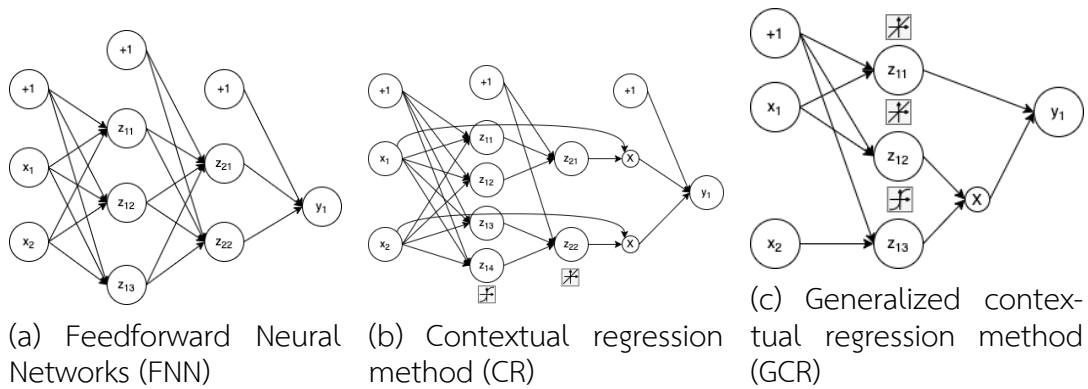
หากเราประมาณ $I(x_2 \geq 0)$ ด้วย tanh function จะได้ว่า

$$I(x_2 \geq 0) \approx \frac{1}{2} (\tanh(\omega x_2) + 1)$$

เมื่อ ω มีค่าสูงพอ ซึ่งเมื่อแทนเข้าไปในสมการที่ 3.9 จะได้ว่า

$$\begin{aligned} y &= I(x_2 \geq 0) \cdot (-73x_1 - 96) + I(x_2 < 0) \cdot (-22x_1 + 35) \\ &= \frac{1}{2} (\tanh(\omega x_2) + 1) \cdot (-73x_1 - 96) + (1 - \frac{1}{2} (\tanh(\omega x_2) + 1)) \cdot (-22x_1 + 35) \\ &= \frac{1}{2} (-95x_1 - 61) + \tanh(\omega x_2) \cdot \frac{1}{2} (-51x_1 - 131) \\ &= z_{11} + z_{12} \cdot z_{13} \end{aligned}$$

ซึ่งจากการจัดรูปข้างต้น จะเห็นได้ว่าสามารถสร้างจัดโครงสร้าง neural network ของกรณีนี้ได้ดังรูปที่ 3.3(c)



รูปที่ 3.3: โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 1

กรณีที่ 2: 3 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)

1. Feedforward Neural Networks (FNN)

สำหรับกรณีนี้ เราต้องการสร้าง Neural network ที่มีความสามารถในการจำแนก 3 บริบท และสามารถสร้างความสัมพันธ์เชิงเส้น 3 รูปแบบ (บริบทละ 1 รูปแบบ) จึงมีการกำหนดให้ชั้นที่ซ่อนอยู่ที่ 1 ประกอบไปด้วย 4 โหนด โดยโหนดที่ 1 มีไว้เพื่อจำแนกบริบทจากข้อมูล, โหนดที่ 2 มีไว้สำหรับการทำนายผลเฉลี่ยของบริบทที่ 1, โหนดที่ 3 มีไว้สำหรับการทำนายผลเฉลี่ยของบริบทที่ 2, โหนดที่ 4 มีไว้สำหรับการทำนายผลเฉลี่ยของบริบทที่ 3 และชั้นที่ซ่อนอยู่ที่ 2 ประกอบไปด้วย 3 โหนด เพื่อเพิ่มความซับซ้อนของตัวแบบ ก่อนที่จะเข้าสู่ชั้นข้อมูลออก โดยจะมีโครงสร้างดังรูปที่ 3.4(a)

2. Contextual regression method (CR)

ในกรณีนี้มีจำนวนตัวแปรเข้าเท่ากับกรณีที่ 1 แต่เนื่องจากเป็นกรณี 3 บริบทที่มีความซับซ้อนกว่า จึงแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบทด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 3 โหนด เพื่อให้มีความซับซ้อนใกล้เคียงกันกับตัวแบบ FNN โดยจะมีโครงสร้างดังรูปที่ 3.4(b)

3. Generalized contextual regression method (GCR)

สืบเนื่องมาจากสมการ 2.2 หากเราจัดรูปใหม่ให้สอดคล้องกับสมการนี้จะได้ว่า

$$y = I(x_2 \geq \frac{1}{3}) \cdot (22x_1 + 60) + I(\frac{1}{3} > x_2 \geq -\frac{1}{3}) \cdot (-47x_1 - 73) + I(x_2 < -\frac{1}{3}) \cdot (60x_1 + 55) \tag{3.10}$$

หากเราประมาณ $I(x_2 \geq \frac{1}{3})$ และ $I(x_2 \geq -\frac{1}{3})$ ด้วย tanh function จะได้ว่า

$$I(x_2 \geq \frac{1}{3}) \approx \frac{1}{2} (\tanh(\omega_1 x_2 - \frac{1}{3}) + 1)$$

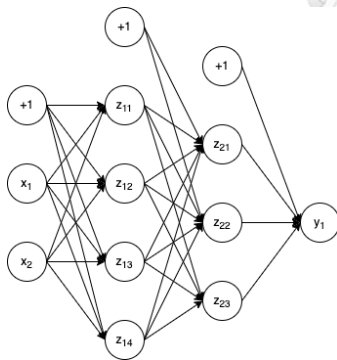
และ

$$I(x_2 \geq -\frac{1}{3}) \approx \frac{1}{2} (\tanh(\omega_2 x_2 + \frac{1}{3}) + 1)$$

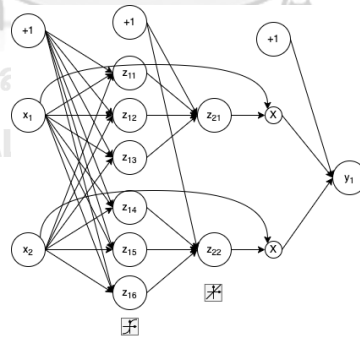
เมื่อ ω_1 และ ω_2 มีค่าสูงพอ ซึ่งเมื่อแทนเข้าไปในสมการที่ 3.10 จะได้ว่า

$$\begin{aligned} y &= I(x_2 \geq \frac{1}{3}) \cdot (22x_1 + 60) + I(\frac{1}{3} > x_2 \geq -\frac{1}{3}) \cdot (-47x_1 - 73) + I(x_2 < -\frac{1}{3}) \cdot (60x_1 + 55) \\ &= I(x_2 \geq \frac{1}{3}) \cdot (22x_1 + 60) + (I(x_2 \geq -\frac{1}{3}) - I(x_2 \geq \frac{1}{3})) \cdot (-47x_1 - 73) \\ &\quad + (1 - I(x_2 \geq -\frac{1}{3})) \cdot (60x_1 + 55) \\ &= 60x_1 + 55 + I(x_2 \geq \frac{1}{3}) \cdot (-25x_1 - 13) + I(x_2 \geq -\frac{1}{3}) \cdot (-107x_1 - 128) \\ &= 60x_1 + 55 + \frac{1}{2} (\tanh(\omega_1 x_2 - \frac{1}{3}) + 1) \cdot (-25x_1 - 13) \\ &\quad + \frac{1}{2} (\tanh(\omega_2 x_2 + \frac{1}{3}) + 1) \cdot (-107x_1 - 128) \\ &= \frac{1}{2} (-12x_1 - 31) + \tanh(\omega_1 x_2 - \frac{1}{3}) \cdot \frac{1}{2} (-25x_1 - 13) + \tanh(\omega_2 x_2 + \frac{1}{3}) \cdot \frac{1}{2} (-107x_1 - 128) \\ &= z_{11} + z_{12} \cdot z_{14} + z_{13} \cdot z_{15} \end{aligned}$$

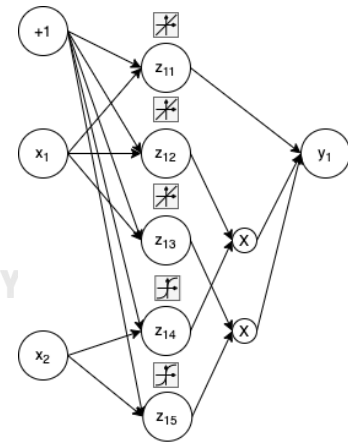
ซึ่งจากการจัดรูปข้างต้น จะเห็นได้ว่าสามารถสร้างจัดโครงสร้าง neural network ของกรณีนี้ได้ ดังรูปที่ 3.4(c)



(a) Feedforward Neural Networks (FNN)



(b) Contextual regression method (CR)



(c) Generalized contextual regression method (GCR)

รูปที่ 3.4: โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 2

กรณีศึกษาที่ 3.1: 2 ชั้น 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)

1. Feedforward Neural Networks (FNN)

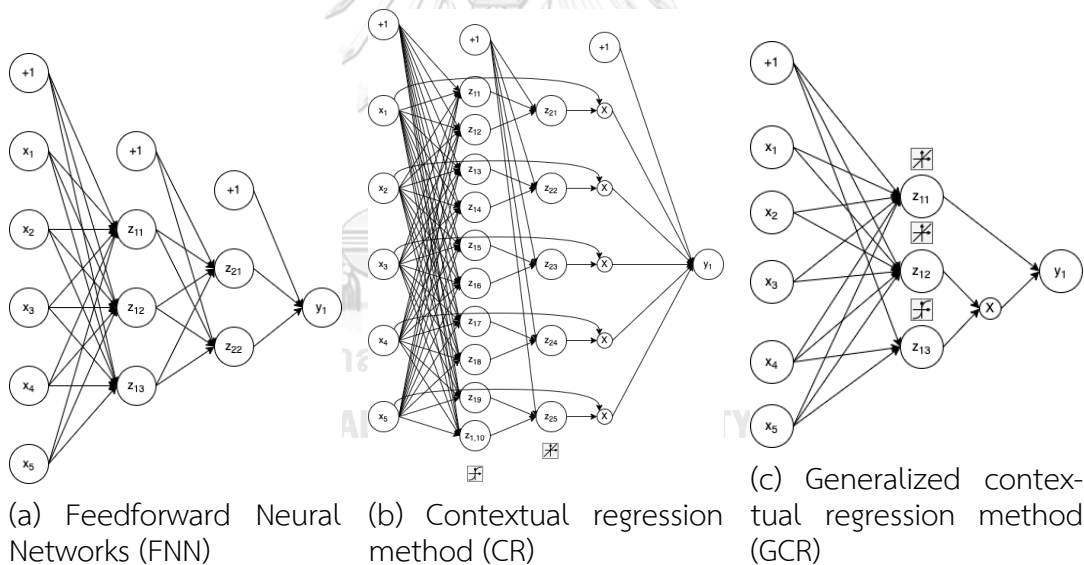
เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 1 แต่แตกต่างกันที่จำนวนตัวแปรเข้า จึงสามารถสร้างโครงสร้างที่มีรูปแบบคล้ายกันได้ดังรูปที่ 3.5(a)

2. Contextual regression method (CR)

ในการทำงานเดียวกับกับตัวแบบ CR ในกรณีศึกษาก่อนหน้านี้ จะแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบทด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 2 โหนด โดยจะมีโครงสร้างดังรูปที่ 3.5(b)

3. Generalized contextual regression method (GCR)

เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 1 แต่แตกต่างกันที่จำนวนตัวแปรเข้า จึงสามารถสร้างโครงสร้างที่มีรูปแบบคล้ายกันได้ดังรูปที่ 3.5(c)



รูปที่ 3.5: โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 3.1

กรณีศึกษาที่ 3.2: 2 ชั้น 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว (มีความสัมพันธ์แบบบริบทไม่เชิงเส้น)

1. Feedforward Neural Networks (FNN)

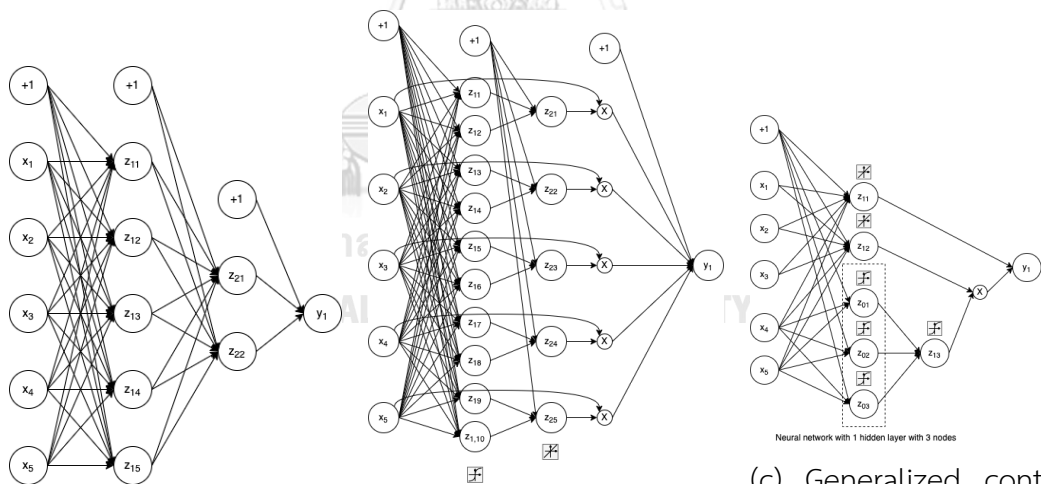
เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 3.1 แต่แตกต่างกันที่มีความซับซ้อนจากการเป็นกรณีบริบทไม่เชิงเส้น จึงเพิ่มความซับซ้อนให้ตัวแบบโดยการเพิ่มจำนวนโหนดในชั้นที่ซ่อนอยู่ที่ 1 เป็น 5 โหนด จึงสามารถสร้างโครงสร้างของตัวแบบในกรณีนี้ได้ดังรูปที่ 3.6(a)

2. Contextual regression method (CR)

ในการทำงานเดียวกับกับตัวแบบ CR ในกรณีศึกษาก่อนหน้านี้ จะแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบทด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 2 โหนด โดยจะมีโครงสร้างดังรูปที่ 3.6(b)

3. Generalized contextual regression method (GCR)

เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 3.1 แต่แตกต่างกันที่กรณีนี้เป็นกรณีบริบทไม่เชิงเส้น จึงเพิ่มความซับซ้อนให้ตัวแบบโดยการเพิ่ม neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 3 โหนด ให้แก่ตัวแปรเชิงบริบทก่อนที่จะส่งค่าไปยังโหนด z_{13} ดังรูปที่ 3.6(c)



(a) Feedforward Neural Networks (FNN)

(b) Contextual regression method (CR)

(c) Generalized contextual regression method (GCR)

รูปที่ 3.6: โครงสร้าง neural network ของตัวแบบ 3 ประเภท สำหรับกรณีศึกษาที่ 3.2

กรณีศึกษาที่ 4.1: 3 ชั้น 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว (มีความสัมพันธ์แบบบริบทเชิงเส้น)

1. Feedforward Neural Networks (FNN)

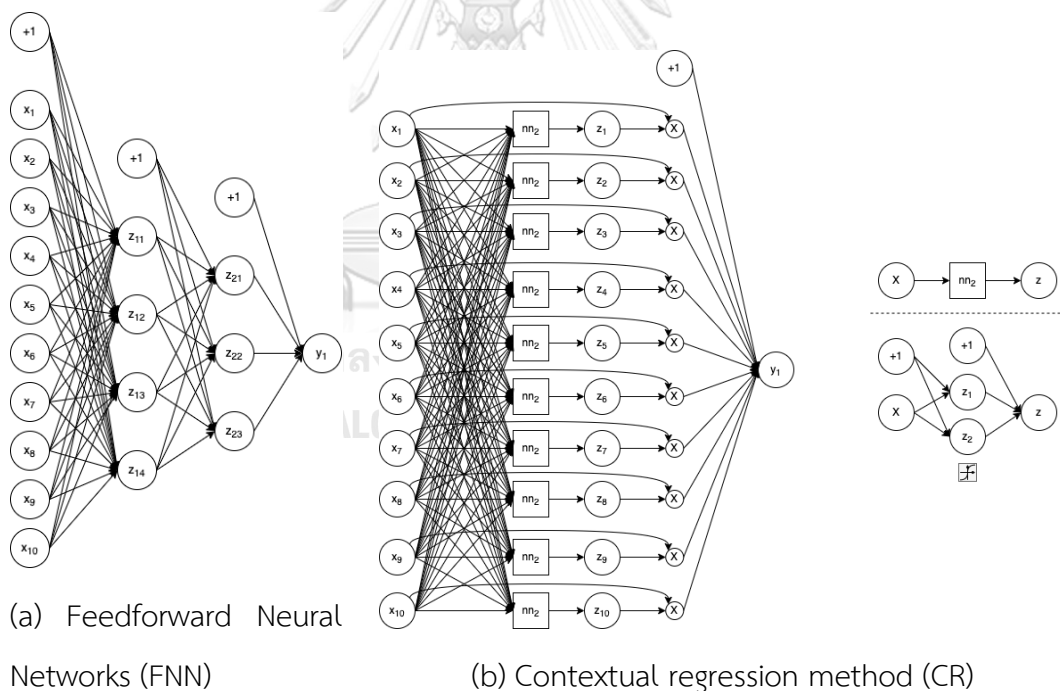
เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 1 แต่แตกต่างกันที่จำนวนตัวแปรเข้า จึงสามารถสร้างโครงสร้างที่มีรูปแบบคล้ายกันได้ดังรูปที่ 3.7(a)

2. Contextual regression method (CR)

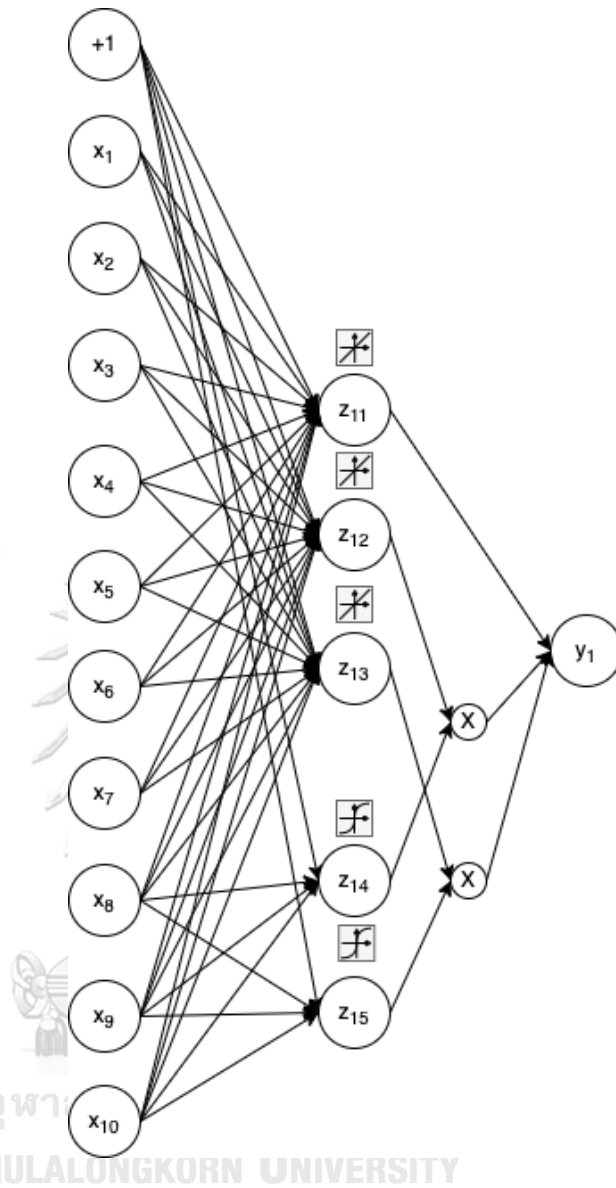
ในการทำงานเดียวกับกับตัวแบบ CR ในกรณีศึกษาก่อนหน้านี้ จะแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบทด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 2 โหนด โดยจะมีโครงสร้างดังรูปที่ 3.7(b)

3. Generalized contextual regression method (GCR)

เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 2 แต่แตกต่างกันที่จำนวนตัวแปรเข้า จึงสามารถสร้างโครงสร้างที่มีรูปแบบคล้ายกันได้ดังรูปที่ 3.8



รูปที่ 3.7: โครงสร้าง neural network ของตัวแบบ FNN และตัวแบบ CR สำหรับกรณีศึกษาที่ 4.1



รูปที่ 3.8: โครงสร้าง neural network ของตัวแบบ GCR สำหรับกรณีศึกษาที่ 4.1

กรณีศึกษาที่ 4.2: 3 ชั้น 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว (มีความสัมพันธ์แบบบริบทไม่เชิงเส้น)

1. Feedforward Neural Networks (FNN)

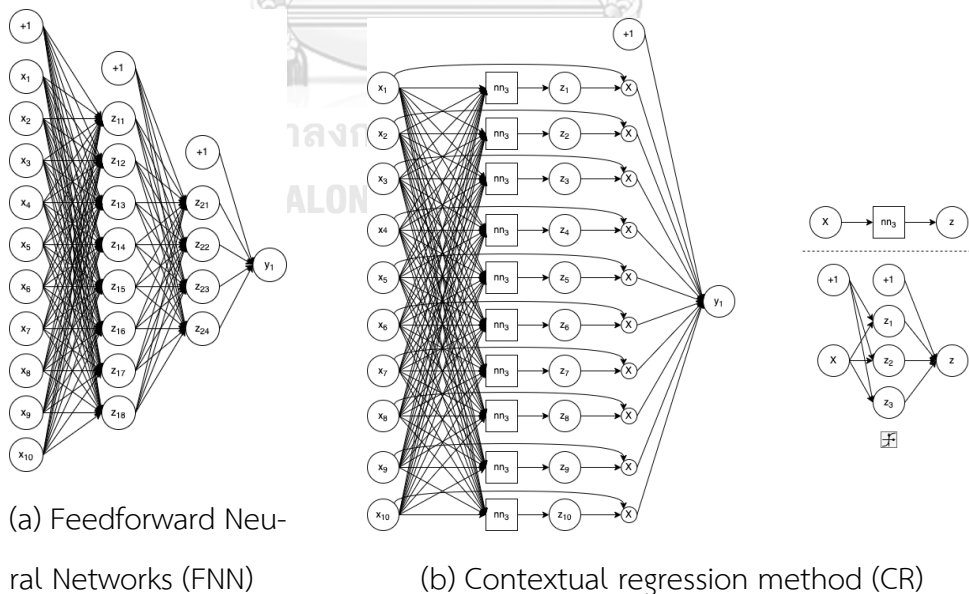
เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 4.1 แต่แตกต่างกันที่มีความซับซ้อนจากการเป็นกรณีบริบทไม่เชิงเส้น จึงเพิ่มความซับซ้อนให้ตัวแบบโดยการเพิ่มจำนวนโหนดในชั้นที่ซ่อนอยู่ที่ 1 เป็น 8 โหนด และจำนวนโหนดในชั้นที่ซ่อนอยู่ที่ 2 เป็น 4 โหนด จึงสามารถสร้างโครงสร้างของตัวแบบในกรณีนี้ได้ดังรูปที่ 3.9(a)

2. Contextual regression method (CR)

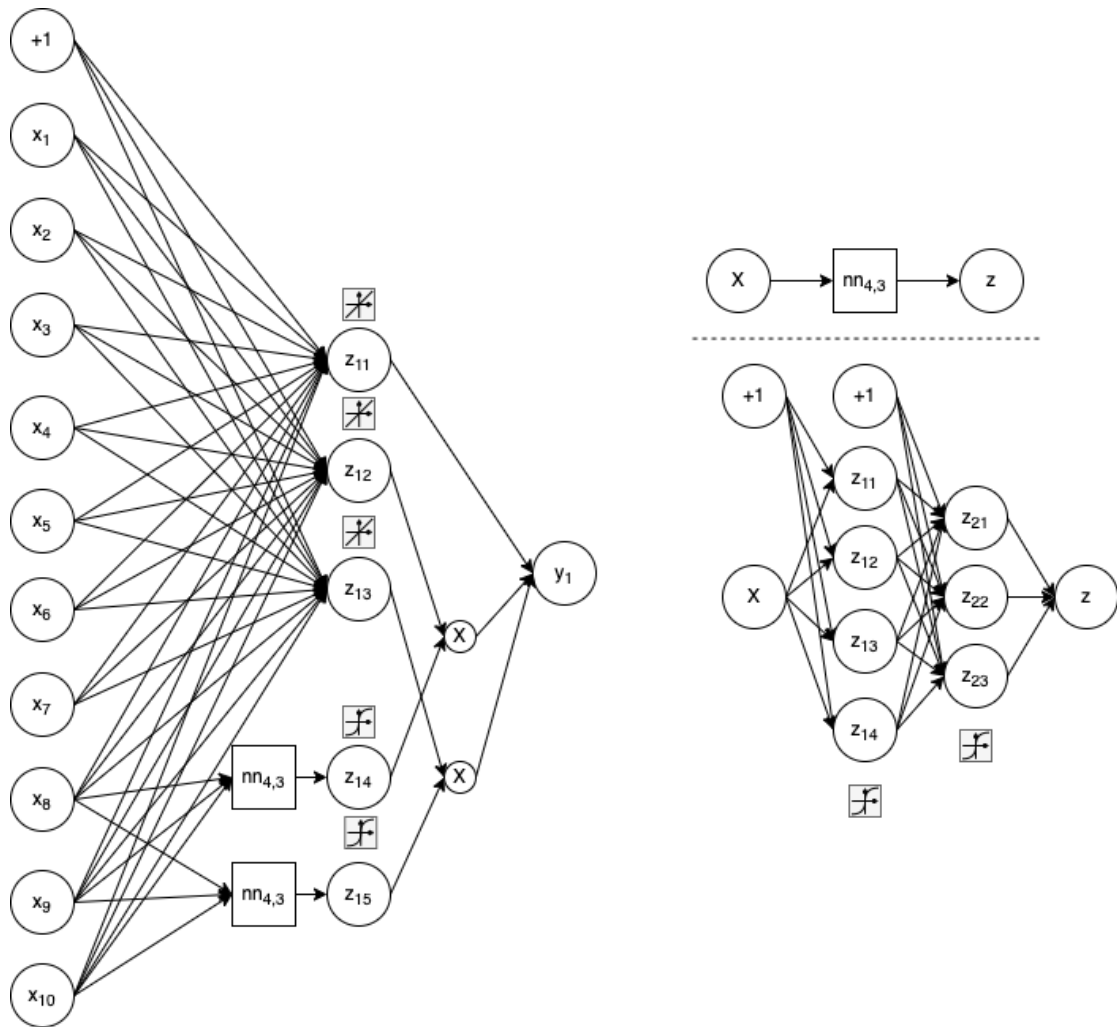
ในการทำงานเดียวกับกับตัวแบบ CR ในกรณีก่อนหน้านี้ จะแทนฟังก์ชันค่าถ่วงน้ำหนักเชิงบริบทด้วย neural network ที่มี 1 ชั้นที่ซ่อนอยู่ขนาด 3 โหนด โดยจะมีโครงสร้างดังรูปที่ 3.9(b)

3. Generalized contextual regression method (GCR)

เนื่องจากกรณีศึกษานี้มีรูปแบบเดียวกับกรณีศึกษาที่ 4.1 แต่แตกต่างกันที่กรณีนี้เป็นกรณีบริบทไม่เชิงเส้น จึงเพิ่มความซับซ้อนให้ตัวแบบโดยการเพิ่ม neural network ที่มี 2 ชั้นที่ซ่อนอยู่ขนาด 4 โหนด และ 3 โหนดตามลำดับ ให้แก่ตัวแปรเชิงบริบทก่อนที่จะส่งค่าไปยังโหนด z_{14} และ z_{15} ดังรูปที่ 3.10



รูปที่ 3.9: โครงสร้าง neural network ของตัวแบบ FNN และตัวแบบ CR สำหรับกรณีศึกษาที่ 4.2



รูปที่ 3.10: โครงสร้าง neural network ของตัวแบบ GCR สำหรับกรณีศึกษาที่ 4.2

3.4 ขั้นตอนในการดำเนินงานวิจัย

สำหรับการวิเคราะห์ข้อมูลแต่ละกรณี มีขั้นตอนในการดำเนินงานวิจัยดังนี้

1. ทำการจำลองข้อมูลแต่ละกรณีตาม 3.1
2. แบ่งข้อมูลที่จำลองออกเป็นข้อมูลสอน 80% และข้อมูลทดสอบ 20%
3. ออกแบบโครงสร้าง FNN, CR และ GCR ที่เหมาะสม โดยให้โครงสร้างมีความซับซ้อนใกล้เคียงกันโดยมีแนวคิดในการสร้างดังที่แสดงไว้ใน 3.3
4. สำหรับแต่ละตัวแบบ จะสุ่มค่าถ่วงน้ำหนักเริ่มต้น ที่ใช้ในการวิเคราะห์ข้อมูล โดยจะใช้วิธี back-propagate ที่กำหนดพารามิเตอร์ต่าง ๆ ตาม 3.2 เพื่อให้ได้มาซึ่งค่าถ่วงน้ำหนักสุดท้ายที่ฟังก์ชันค่าสูญเสียให้ค่าต่ำสุด
5. สำหรับแต่ละตัวแบบ นำค่าถ่วงน้ำหนักสุดท้ายที่ได้ มาใช้คำนวณผลลัพธ์ทำนาย จากการวิเคราะห์ข้อมูลทดสอบ และทำการคำนวณประสิทธิภาพเทียบกับผลลัพธ์จริง ด้วยเกณฑ์วัดดังนี้

- 5.1. ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean squared error; MSE)

$$MSE = \|\hat{y} - y\|^2$$

เมื่อ

y เป็นเวกเตอร์ผลลัพธ์จริง และ

\hat{y} เป็นเวกเตอร์ผลลัพธ์ทำนาย

- 5.2. ค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of determination; R^2)

$$R^2 = 1 - \frac{\|\hat{y} - y\|^2}{\|\bar{y} - y\|^2}$$

เมื่อ

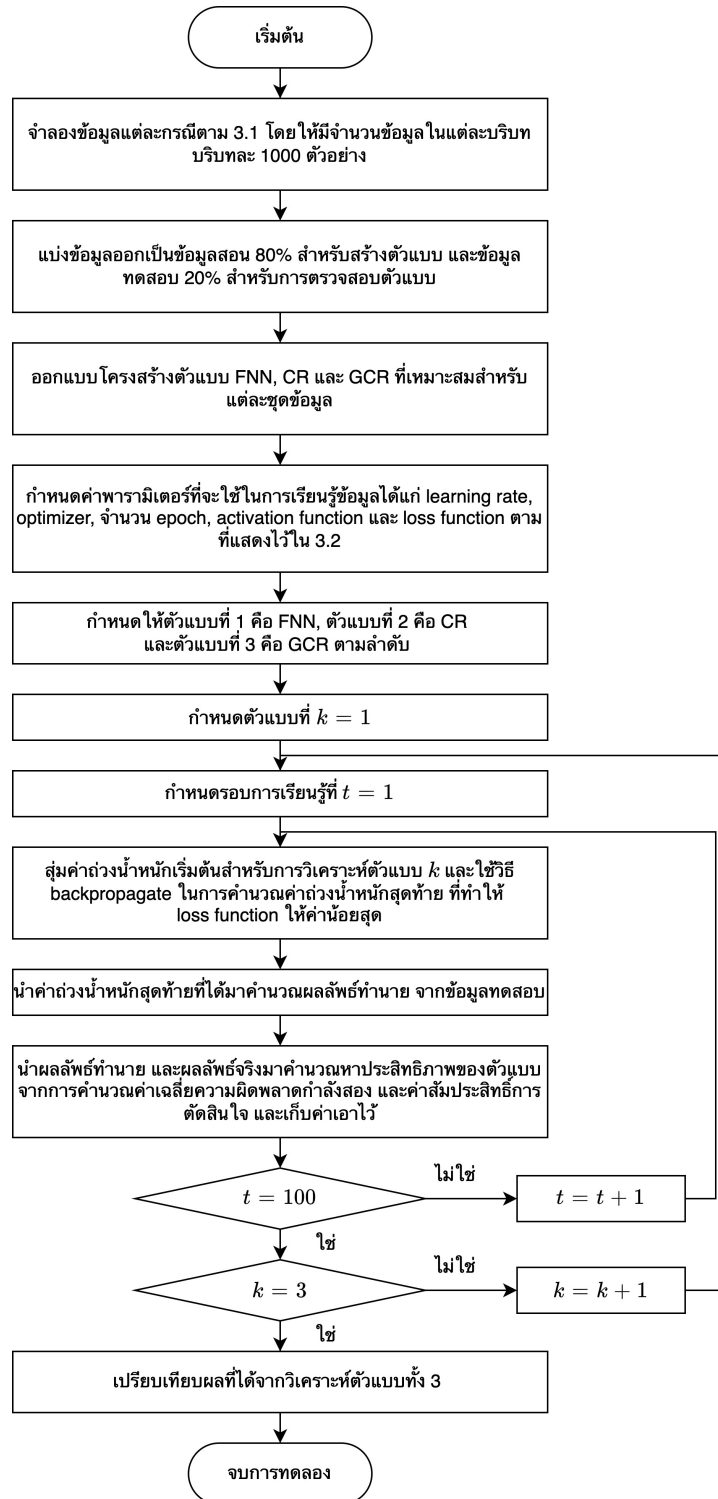
y เป็นเวกเตอร์ผลลัพธ์จริง,

\hat{y} เป็นเวกเตอร์ผลลัพธ์ทำนาย และ

\bar{y} เป็นเวกเตอร์ค่าเฉลี่ยของผลลัพธ์จริงทั้งหมด

6. ทำซ้ำขั้นตอน 4 และ 5 ด้วยการสุ่มค่าถ่วงน้ำหนักเริ่มต้น 100 รอบต่อ 1 ตัวแบบ ต่อ 1 กรณีศึกษา และรวบรวมผลลัพธ์ทั้งหมด
7. สรุปผลการทดลอง

3.5 แผนผังแสดงขั้นตอนการทำงานในแต่ละกรณีทดลอง



บทที่ 4

ผลการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบโครงสร้างตัวแบบ GCR ที่ผู้วิจัยพัฒนาขึ้นมา กับตัวแบบ FNN และ CR โดยผ่านการจำลองข้อมูลที่มีความสัมพันธ์เชิงบริบทในรูปแบบเชิงเส้น และไม่เชิงเส้น ที่กล่าวไว้ในบทที่ 3 ดังนั้นผลการวิจัยจะแบ่งออกเป็น 5 ส่วนดังนี้

1. กรณี 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น
2. กรณี 3 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น
3. กรณี 2 บริบท 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว
 - a) ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น
 - b) ข้อมูลมีความสัมพันธ์แบบบริบทไม่เชิงเส้น
4. กรณี 3 บริบท 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว
 - a) ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น
 - b) ข้อมูลมีความสัมพันธ์แบบบริบทไม่เชิงเส้น
5. ผลการวิเคราะห์จุดแบ่งระหว่างบริบทสำหรับตัวแบบ GCR

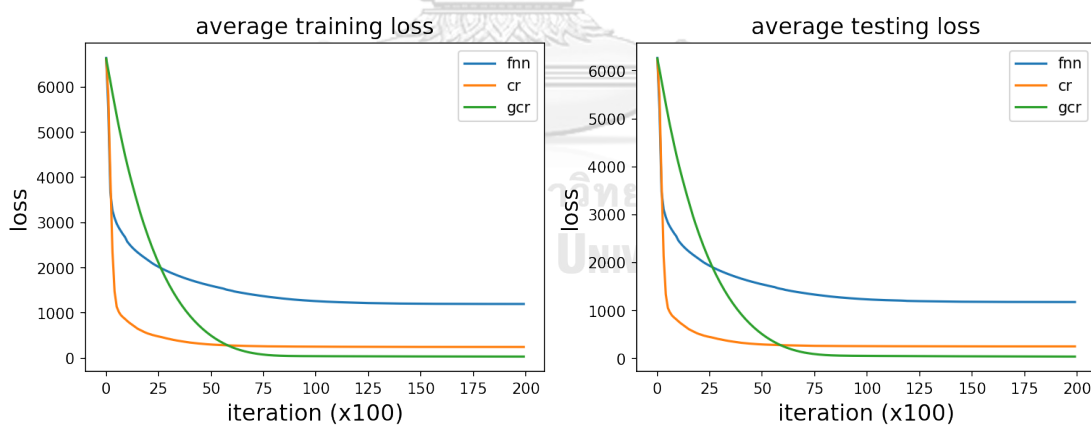
4.1 กรณี 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.1

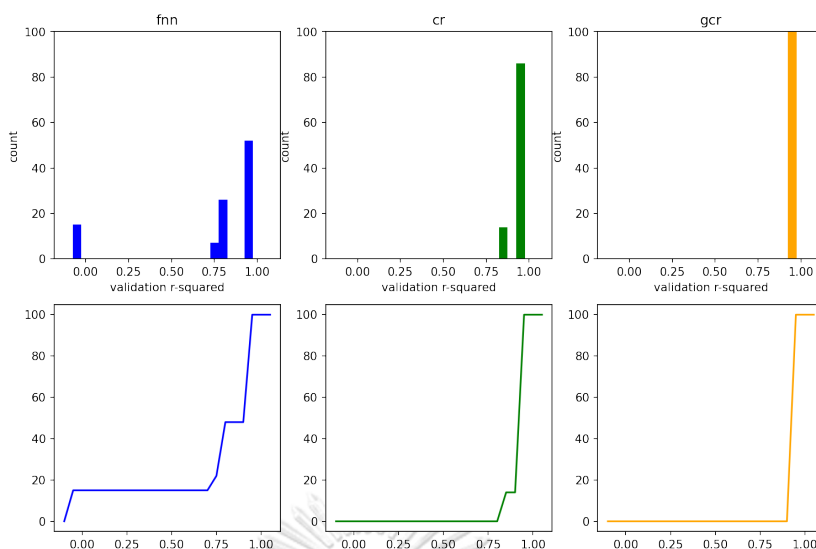
ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	20	22	6
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	22	25	11
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	1179.72	254.44	41.98
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	27.58	159.51	40.21
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.7815	0.9529	0.9922
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.9949	0.9705	0.9926

ตารางที่ 4.1: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.1

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.1 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.2



รูปที่ 4.1: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบ สำหรับแต่ละตัวแบบ



รูปที่ 4.2: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ
(ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่าในกรณีนี้ หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ FNN ดีกว่า GCR ดีกว่า CR แต่หากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า CR ดีกว่า FNN โดยเวลาที่ใช้การเรียนรู้ตัวแบบ จะขึ้นกับจำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) ซึ่งจะได้ว่า GCR ใช้เวลาน้อยกว่า FNN ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.1 จะเห็นว่าโดยเฉลี่ยแล้วตัวแบบ FNN และ CR ใช้เวลาเร็วกว่าตัวแบบ GCR และมีประสิทธิภาพที่ดีกว่าในช่วง epoch เริ่มต้น แต่เมื่อพิจารณาที่ epoch สุดท้าย จะพบว่าโดยเฉลี่ยแล้วตัวแบบ GCR ให้ประสิทธิภาพที่ดีกว่า CR และ FNN ตามลำดับ และจากรูปที่ 4.2 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 โดยส่วนใหญ่ เมื่อเทียบกับตัวแบบ FNN และ CR ที่มีโอกาสที่ค่าสัมประสิทธิ์การตัดสินใจมีค่าน้อยกว่า 1 อย่างเห็นได้ชัด

4.2 กรณี 3 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว

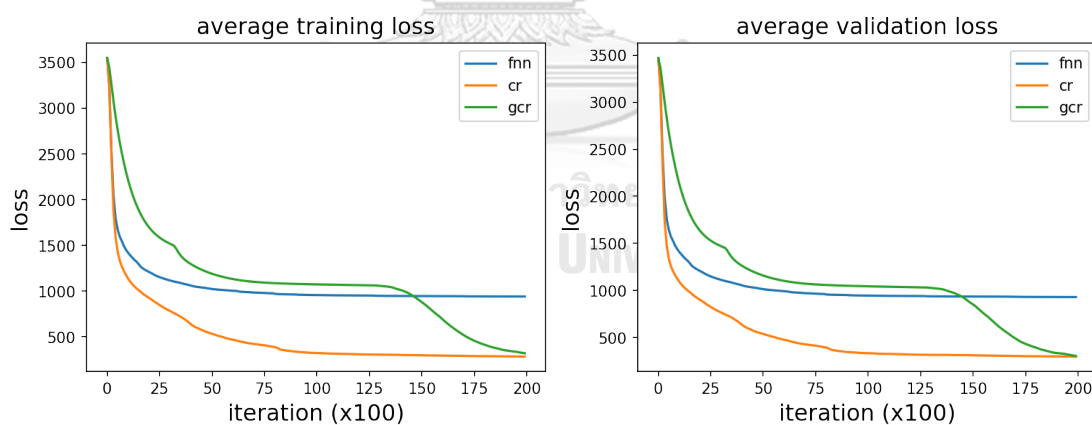
ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.2

ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	31	30	10
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	37	44	20
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	930.46	298.05	301.92
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	25.62	64.02	29.98
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.7312	0.9139	0.9128
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.9926	0.9815	0.9913

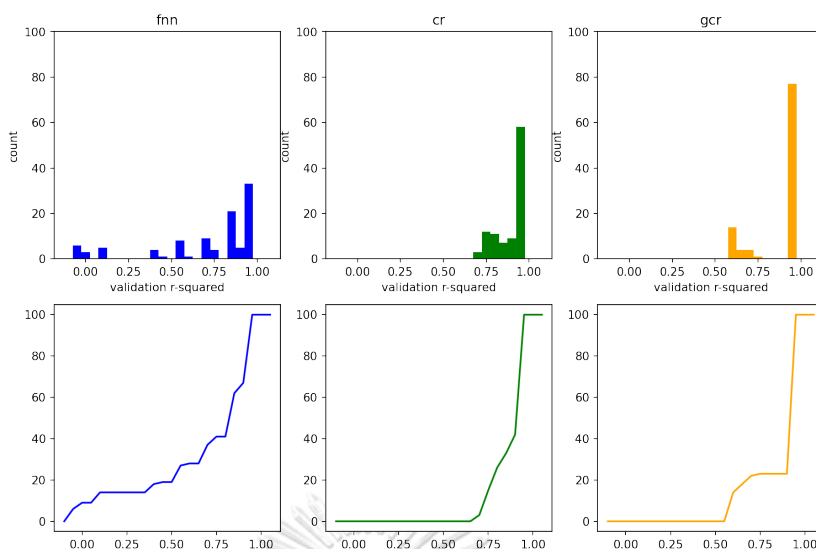
ตารางที่ 4.2: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.2

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.3 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.4



รูปที่ 4.3: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบสำหรับแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่า หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ FNN ดีกว่า GCR ดีกว่า CR แต่หากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า CR ดีกว่า FNN โดยเวลาที่ใช้การเรียนรู้ตัวแบบ จะขึ้นกับ



รูปที่ 4.4: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ (ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) ซึ่งจะได้ว่า GCR ใช้เวลาน้อยกว่า FNN ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.3 จะเห็นว่าโดยเฉลี่ยแล้วตัวแบบ FNN และ CR ใช้เวลาลู่ว่าน้อยกว่าตัวแบบ GCR และมีประสิทธิภาพที่ดีกว่าในช่วง epoch เริ่มต้น แต่เมื่อพิจารณาที่ epoch สุดท้าย จะพบว่าโดยเฉลี่ยแล้วตัวแบบ GCR และ CR ให้ประสิทธิภาพที่ใกล้เคียงกันและดีกว่าตัวแบบ FNN และจากรูปที่ 4.4 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 ในสัดส่วนที่มากกว่าตัวแบบ CR และ FNN ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.3 กรณี 2 บริบท 5 ตัวแปร ตัวแปรเชิงบริบท 2 ตัว

ในกรณีนี้จะแบ่งเป็น 2 กรณีย่อยตามรูปแบบความสัมพันธ์เชิงบริบท ดังนี้

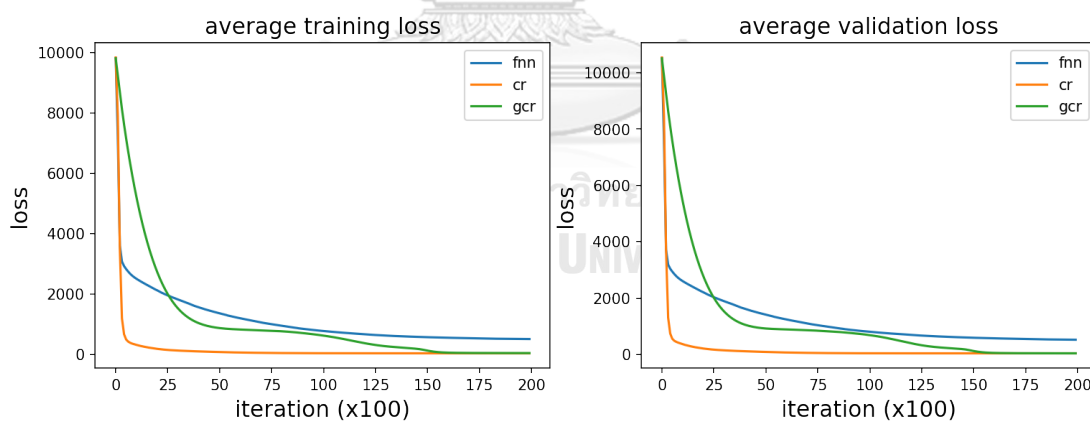
4.3.1 ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.3

ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	29	81	15
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	27	87	19
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	520.39	38.63	39.04
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	27.03	29.03	27.42
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.9249	0.9944	0.9944
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.9961	0.9958	0.9960

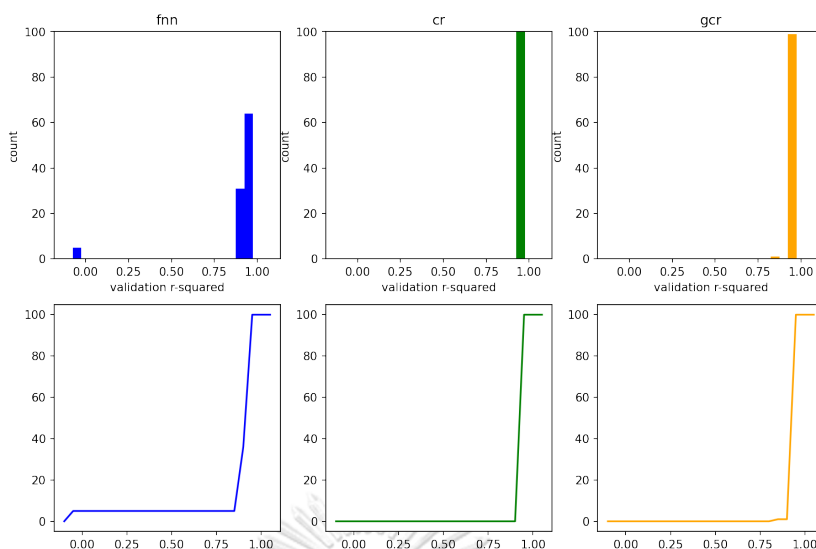
ตารางที่ 4.3: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.3

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.5 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.6



รูปที่ 4.5: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบสำหรับแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่า หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ FNN ดีกว่า GCR ดีกว่า CR แต่หากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ CR ดีกว่า GCR ดีกว่า FNN โดยเวลาที่ใช้การเรียนรู้ตัวแบบ จะขึ้นกับ



รูปที่ 4.6: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ (ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) ซึ่งจะได้ว่า GCR ใช้เวลาน้อยกว่า FNN ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.5 จะเห็นว่าโดยเฉลี่ยแล้วตัวแบบ FNN และ CR ใช้เวลาลู่เข้าน้อยกว่าตัวแบบ GCR และมีประสิทธิภาพที่ดีกว่าในช่วง epoch เริ่มต้น แต่เมื่อพิจารณาที่ epoch สุดท้าย จะพบว่าโดยเฉลี่ยแล้วตัวแบบ GCR และ CR ให้ประสิทธิภาพที่ใกล้เคียงกันและดีกว่าตัวแบบ FNN และจากรูปที่ 4.6 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR และ CR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 ในสัดส่วนที่มากกว่าตัวแบบ FNN

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

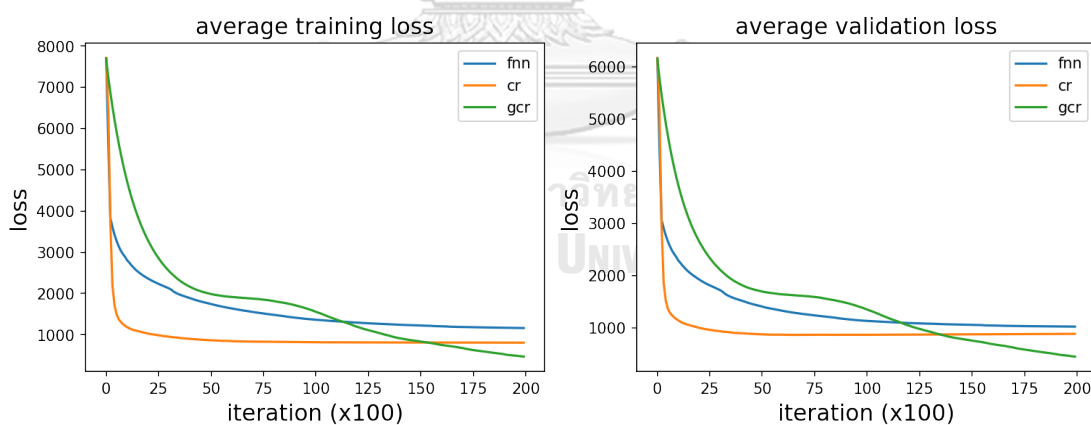
4.3.2 ข้อมูลมีความสัมพันธ์แบบปรียบไม่เชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.4

ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	45	81	25
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	29	80	27
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	1023.10	884.64	446.33
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	447.37	450.06	65.73
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.8148	0.8399	0.9192
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.9136	0.9185	0.9881

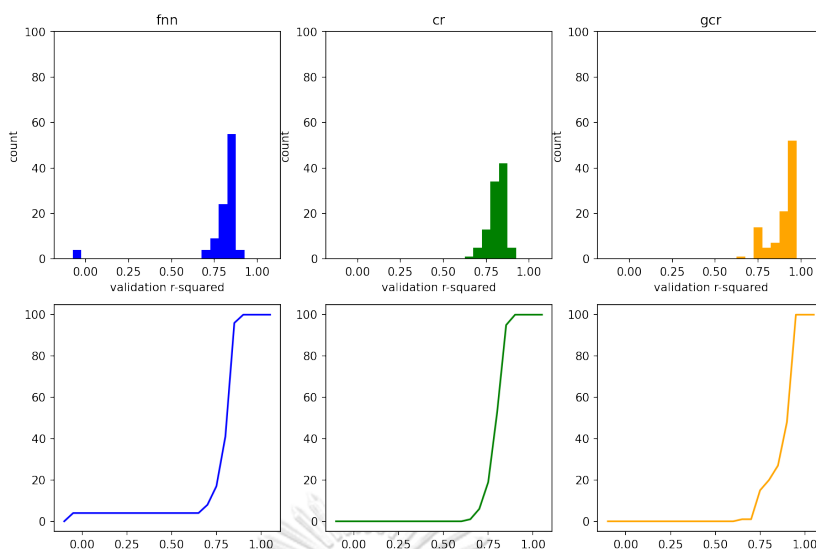
ตารางที่ 4.4: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.3

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.7 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.8



รูปที่ 4.7: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบสำหรับแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่า หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า FNN ดีกว่า CR และหากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า CR ดีกว่า FNN โดยเวลาที่ใช้การเรียนรู้ตัวแบบ จะขึ้นกับ



รูปที่ 4.8: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ (ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) ซึ่งจะได้ว่า GCR ใช้เวลาน้อยกว่า FNN ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.7 จะเห็นว่าโดยเฉลี่ยแล้วตัวแบบ FNN และ CR ใช้เวลาสูงเข้าน้อยกว่าตัวแบบ GCR และมีประสิทธิภาพที่ดีกว่าในช่วง epoch เริ่มต้น แต่เมื่อพิจารณาที่ epoch สุดท้าย จะพบว่าโดยเฉลี่ยแล้วตัวแบบ GCR ให้ประสิทธิภาพที่ดีกว่าตัวแบบ CR ดีกว่าตัวแบบ FNN และจากรูปที่ 4.8 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 ในสัดส่วนที่มากกว่าตัวแบบ CR และตัวแบบ FNN ตามลำดับ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4.4 กรณี 3 บริบท 10 ตัวแปร ตัวแปรเชิงบริบท 3 ตัว

ในกรณีนี้จะแบ่งเป็น 2 กรณีย่อยตามรูปแบบความสัมพันธ์เชิงบริบท ดังนี้

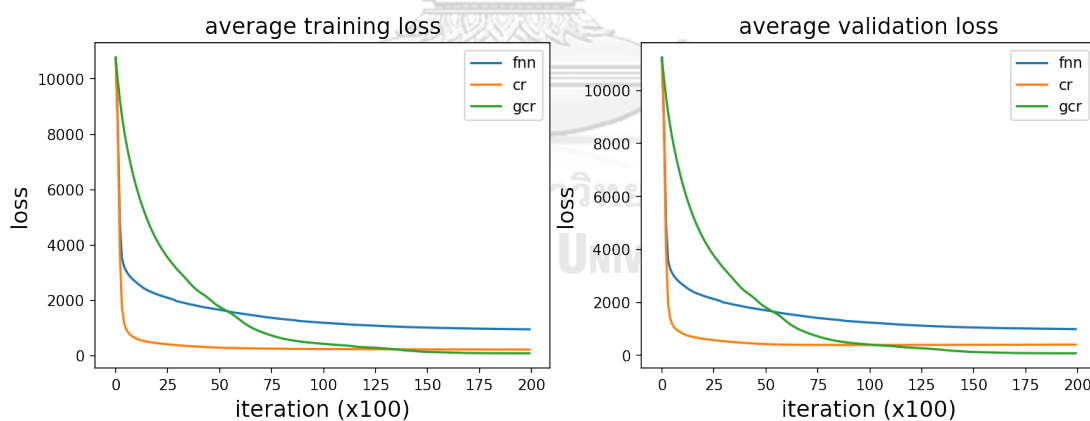
4.4.1 ข้อมูลมีความสัมพันธ์แบบบริบทเชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.5

ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	63	261	41
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	44	232	26
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	988.21	405.02	74.44
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	153.06	178.98	30.22
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.9118	0.9638	0.9934
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.9863	0.9840	0.9973

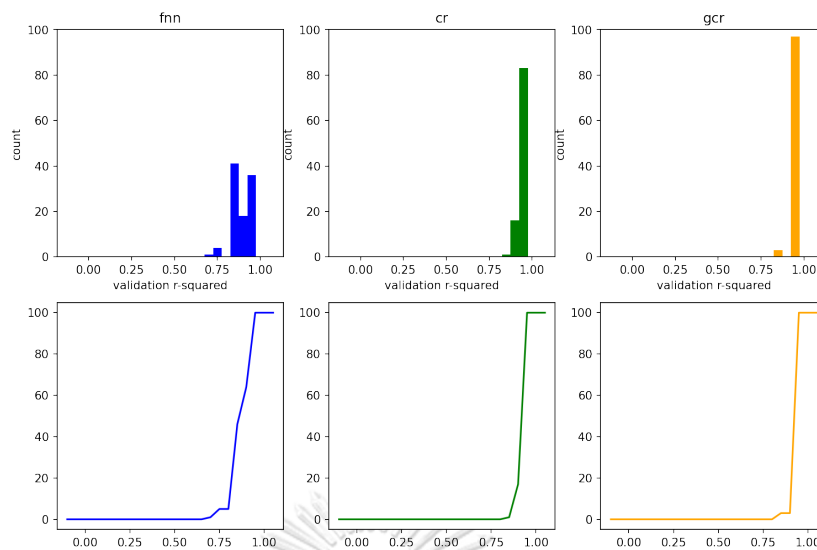
ตารางที่ 4.5: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.4

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.9 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.10



รูปที่ 4.9: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบสำหรับแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่า หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า FNN ดีกว่า CR และหากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า CR ดีกว่า FNN โดยเวลาที่ใช้การเรียนรู้ตัวแบบ จะขึ้นกับ



รูปที่ 4.10: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ (ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) ซึ่งจะได้ว่า GCR ใช้เวลาน้อยกว่า FNN ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.9 จะเห็นว่าโดยเฉลี่ยแล้วตัวแบบ FNN และ CR ใช้เวลาสู่น้อยกว่าตัวแบบ GCR และมีประสิทธิภาพที่ดีกว่าในช่วง epoch เริ่มต้น แต่เมื่อพิจารณาที่ epoch สุดท้าย จะพบว่าโดยเฉลี่ยแล้วตัวแบบ GCR และ CR ให้ประสิทธิภาพที่ใกล้เคียงกันและดีกว่าตัวแบบ FNN และจากรูปที่ 4.10 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 ในสัดส่วนที่มากกว่าตัวแบบ CR และตัวแบบ FNN ตามลำดับ

CHULALONGKORN UNIVERSITY

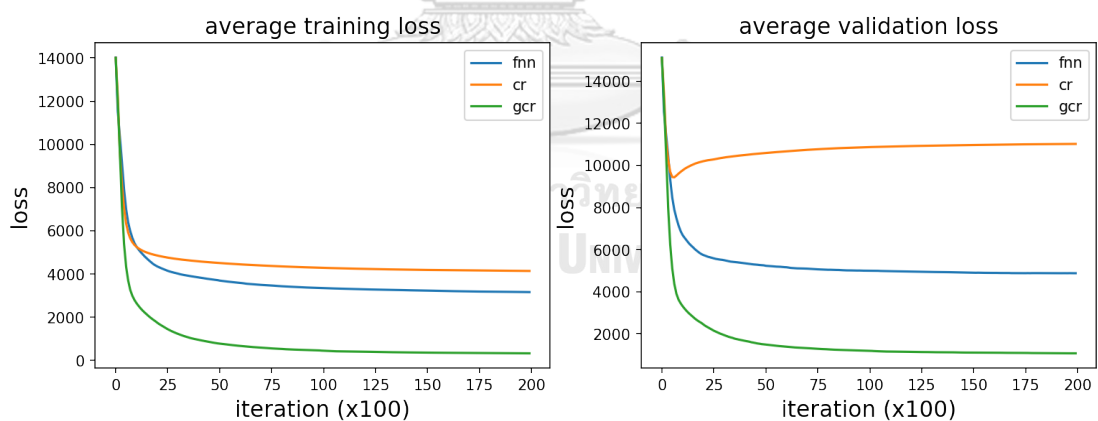
4.4.2 ข้อมูลมีความสัมพันธ์แบบปรียบไม่เชิงเส้น

จากการทดลอง ค่าเฉลี่ยของประสิทธิภาพของแต่ละตัวแบบที่วัดจากข้อมูลทดสอบ จากการทดลอง 100 รอบ แสดงดังตาราง 4.6

ตัวแบบ	FNN	CR	GCR
จำนวนพารามิเตอร์ที่ใช้	129	381	103
เวลาที่ใช้ในการเรียนรู้ตัวแบบ (วินาทีต่อรอบ)	44	251	69
ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง	4875.24	11022.39	1074.76
ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง	2896.02	8619.90	441.74
ค่าเฉลี่ยของค่าสัมประสิทธิ์การตัดสินใจ	0.6685	0.2504	0.9269
ค่าสูงสุดของค่าสัมประสิทธิ์การตัดสินใจ	0.8031	0.4138	0.9700

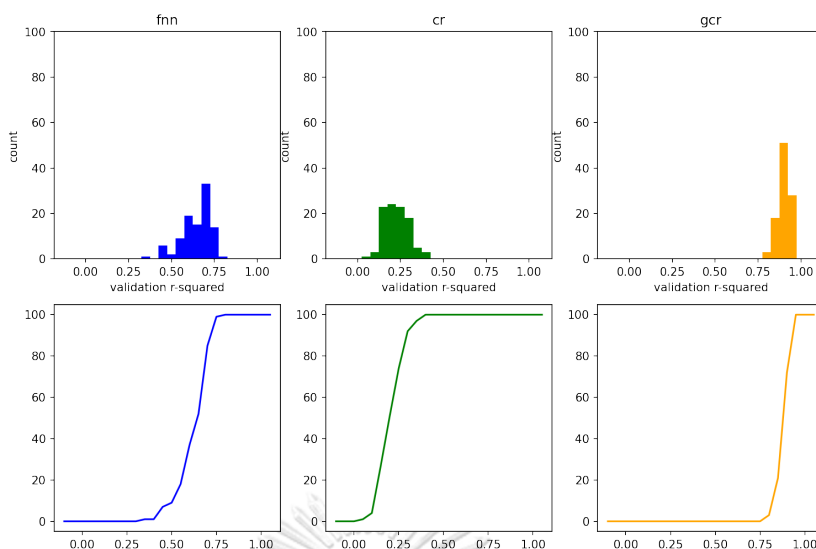
ตารางที่ 4.6: ตารางแสดงจำนวนพารามิเตอร์ของตัวแบบ, เวลาที่ใช้ในการเรียนรู้ของแต่ละตัวแบบ และผลลัพธ์ที่วัดจากการนำข้อมูลทดสอบมาวิเคราะห์ด้วยตัวแบบที่ได้ในกรณี 4.4

โดยมีค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสองในแต่ละ epoch ที่ทำการทดลองตัวแบบดังรูปที่ 4.11 และค่าสัมประสิทธิ์การตัดสินใจจากการทดลอง 100 รอบ ของแต่ละตัวแบบมีการกระจายตัวของดังรูปที่ 4.12



รูปที่ 4.11: กราฟแสดงค่าเฉลี่ย MSE (loss) ในแต่ละ epoch ของข้อมูลสอน และข้อมูลทดสอบสำหรับแต่ละตัวแบบ

จากผลลัพธ์ที่ได้ สามารถสรุปได้ว่า หากพิจารณาจากค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า FNN ดีกว่า CR และหากพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง จะได้ว่าตัวแบบ GCR ดีกว่า FNN ดีกว่า CR เช่นเดียวกัน โดยเวลาที่ใช้การเรียนรู้ตัวแบบ



รูปที่ 4.12: (บน) กราฟแสดงการกระจายตัวของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ (ล่าง) กราฟแสดงการกระจายตัวสะสมของสัมประสิทธิ์การตัดสินใจของแต่ละตัวแบบ

จะขึ้นกับจำนวนพารามิเตอร์ที่ใช้ (ขนาดของ weight) และความซับซ้อนเชิงโครงสร้างของตัวแบบ ซึ่งจะได้ว่า FNN ใช้เวลาน้อยกว่า GCR ใช้เวลาน้อยกว่า CR เมื่อวิเคราะห์ในแต่ละ epoch จากรูปที่ 4.11 เห็นได้ชัดว่าโดยเฉลี่ยแล้วตัวแบบ GCR ทุ่มเข้าสู่ค่าความผิดพลาดที่น้อยกว่าตัวแบบ FNN และ CR ตามลำดับ ในขณะที่ตัวแบบ CR เกิดการแก้ไขปัญหาเกินพอดี (overfitting) เพราะมีความแตกต่างระหว่างความผิดพลาดกำลังสองของผลลัพธ์จากข้อมูลสอน และผลลัพธ์จากข้อมูลทดสอบ อย่างเห็นได้ชัด และจากรูปที่ 4.12 เมื่อพิจารณาการกระจายตัว และการกระจายตัวสะสมของค่าสัมประสิทธิ์การตัดสินใจ พบว่าตัวแบบ GCR ให้ค่าสัมประสิทธิ์การตัดสินใจอยู่ใกล้กับ 1 ในสัดส่วนที่มากกว่าตัวแบบ FNN และตัวแบบ CR ตามลำดับ

4.5 ผลการวิเคราะห์จุดแบ่งระหว่างบริบทสำหรับตัวแบบ GCR

ในที่นี้จะกล่าวถึงผลการวิเคราะห์จุดแบ่งระหว่างบริบทเมื่อใช้ตัวแบบ GCR กับข้อมูลในกรณีที่ 1 เท่านั้น (กรณี 2 บริบท 2 ตัวแปร ตัวแปรเชิงบริบท 1 ตัว) เพราะเป็นกรณีที่เห็นผลได้ชัดเจนที่สุด โดยมีผลลัพธ์ดังนี้

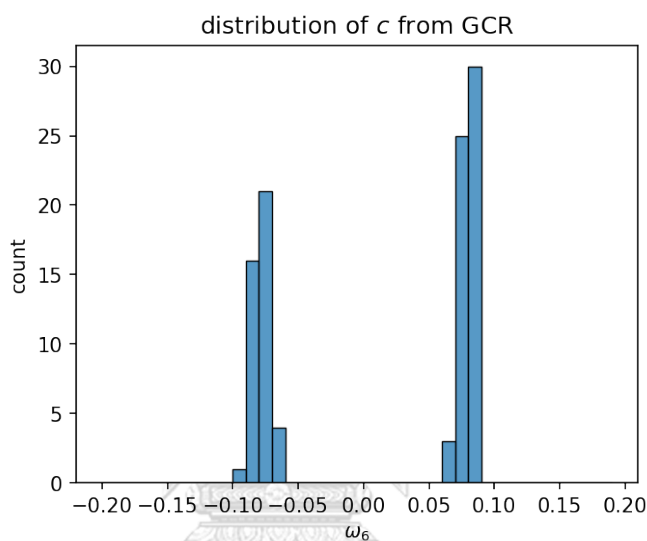
จากข้อมูลที่จำลองในกรณีดังกล่าวสอดคล้องกับสมการ 3.2

$$y = \begin{cases} -73x_1 - 96 + \epsilon_1 & \text{เมื่อ } x_2 \geq 0 \\ -22x_1 + 35 + \epsilon_2 & \text{เมื่อ } x_2 < 0 \end{cases}$$

ประกอบกับโครงสร้างของ GCR จากรูป 3.3(c) ที่หากกำหนดให้ค่าถ่วงน้ำหนักเป็น $\omega = [\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6]^t$ จะได้ว่าโครงสร้างดังกล่าวจะสามารถเขียนสมการความสัมพันธ์ออกมาได้เป็น

$$y = \omega_1 x_1 + \omega_2 + (\omega_3 x_1 + \omega_4) \cdot \tanh(\omega_5 x_2 + \omega_6) \quad (4.1)$$

ซึ่งเห็นได้ว่าจุดแบ่งระหว่างบริบทในกรณีนี้คือ 0 ซึ่งจะตรงกับค่า ω_6 ในสมการ 4.1 ทั้งนี้เมื่อเราทำการวิเคราะห์ค่า ω_6 จากการทดลอง 100 รอบจะได้รับการกระจายตัวดังรูป 4.13 ซึ่งจะเห็นได้ว่าจุดแบ่งที่



รูปที่ 4.13: กราฟแสดงการกระจายตัวของค่า ω_6

ได้จากค่าถ่วงน้ำหนักสุดท้ายอยู่ในช่วง $[-0.1, 0.1]$ ซึ่งใกล้เคียงกับค่าที่กำหนดไว้ โดยมีค่าเฉลี่ยความผิดพลาดกำลังสองของ ω_6 อยู่ที่ 0.062 นั่นคือนอกจากตัวแบบ GCR สามารถแก้ไขปัญหาค่าความสัมพันธ์เชิงบริบทได้แล้ว ยังสามารถอธิบายได้ถึงจุดแบ่งระหว่างบริบท ซึ่งเป็นองค์ประกอบสำคัญได้อีกด้วย

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ตามวัตถุประสงค์ของงานวิจัยนี้ที่ต้องการเปรียบเทียบโครงสร้างตัวแบบ GCR ที่ผู้วิจัยพัฒนาขึ้นมา กับตัวแบบ FNN และ CR จากผลการวิจัยที่ได้ สามารถสรุปผลลัพธ์ของแต่ละกรณีได้ดังนี้

1. จากผลลัพธ์ 4.1 ซึ่งได้จากการจำลองข้อมูลที่มีความซับซ้อนน้อยที่สุด ถึงแม้ว่าตัวแบบ FNN จะเกิดค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง ที่ดีกว่าตัวแบบ FNN และ GCR แต่เมื่อพิจารณาโดยเฉลี่ยแล้ว พบว่าตัวแบบ GCR ให้ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง ที่ดีกว่าตัวแบบ FNN และ CR รวมทั้งระยะเวลาในการเรียนรู้ตัวแบบ และโอกาสที่จะได้ประสิทธิภาพที่ดีที่สุดก็ดีกว่าเช่นกัน
2. จากผลลัพธ์ 4.2 ซึ่งได้จากการจำลองข้อมูลที่คล้ายกับการวิเคราะห์กรณี 4.1 ต่างกันที่ข้อมูลในกรณีนี้มี 3 บริบท ซึ่งจากผลลัพธ์ที่ได้มีความคล้ายเคียงกันจึงสรุปได้คล้ายกันคือตัวแบบ FNN จะเกิดค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสอง ที่ดีกว่าตัวแบบ FNN และ GCR แต่โดยเฉลี่ยแล้วตัวแบบ GCR และ CR ให้ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง ที่ดีกว่าตัวแบบ FNN โดยตัวแบบ CR ให้ค่าที่ดีกว่าเล็กน้อย แต่แลกมากับการที่ตัวแบบ CR ใช้เวลาการเรียนรู้มากกว่าตัวแบบ FNN และ GCR ตามลำดับ
3. จากผลลัพธ์ 4.3 ซึ่งได้จากการจำลองข้อมูลที่คล้ายกับการวิเคราะห์กรณี 4.1 ต่างกันที่ข้อมูลในกรณีนี้มีจำนวนตัวแปรเข้า และตัวแปรเชิงบริบทที่มากกว่า ซึ่งจากผลลัพธ์ที่ได้จะเห็นได้ว่าตัวแบบ CR และ GCR ได้ผลลัพธ์ที่ใกล้เคียงกันอย่างมาก แต่มีความต่างอย่างเห็นได้ชัดในส่วนของจำนวนพารามิเตอร์ที่ใช้ และเวลาที่ใช้ในการเรียนรู้ตัวแบบ ซึ่งตัวแบบ GCR ใช้ทรัพยากรที่น้อยกว่า ในขณะที่ตัวแบบ FNN แม้จะให้ค่าต่ำสุดของค่าเฉลี่ยความผิดพลาดกำลังสองใกล้เคียงกับตัวแบบทั้ง 2 ข้างต้น แต่โอกาสในการเกิดนั้นน้อยกว่าอย่างเห็นได้ชัด โดยสังเกตจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง
4. จากผลลัพธ์ 4.3 ซึ่งได้จากการจำลองข้อมูลที่คล้ายกับการวิเคราะห์กรณี 4.3 ต่างกันที่ข้อมูลในกรณีนี้พิจารณาความสัมพันธ์ตามบริบทไม่เชิงเส้น ซึ่งจากผลลัพธ์ที่ได้จะเห็นได้ว่าตัวแบบ FNN

และ CR ให้ผลลัพธ์ที่ใกล้เคียงกันอย่างมาก แต่มีความต่างอย่างเห็นได้ชัดในส่วนของจำนวนพารามิเตอร์ที่ใช้ และเวลาที่ใช้ในการเรียนรู้ตัวแบบ ซึ่งตัวแบบ FNN ใช้ทรัพยากรที่น้อยกว่า ในขณะที่ตัวแบบ GCR ให้ประสิทธิภาพที่ดีกว่าตัวแบบ FNN และ CR ทั้งในแง่ของการใช้ทรัพยากร และผลลัพธ์จากค่าเฉลี่ยความผิดพลาดกำลังสอง

5. จากผลลัพธ์ 4.4 ซึ่งได้จากการจำลองข้อมูลที่คล้ายกับการวิเคราะห์กรณี 4.3 ต่างกันที่ข้อมูลในกรณีนี้มีจำนวนตัวแปรเข้า และตัวแปรเชิงบริบทที่มากกว่า เมื่อพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง พบว่าตัวแบบ GCR มีประสิทธิภาพที่ดีที่สุด ตามด้วยตัวแบบ CR และ FNN ตามลำดับ แต่หากพิจารณาที่จำนวนพารามิเตอร์ที่ใช้ และระยะเวลาในการเรียนรู้ตัวแบบ พบว่าตัวแบบ CR ใช้ทรัพยากรน้อยกว่าตัวแบบ FNN และ GCR อย่างมาก
6. จากผลลัพธ์ 4.4 ซึ่งได้จากการจำลองข้อมูลที่คล้ายกับการวิเคราะห์กรณี 4.4 ต่างกันที่ข้อมูลในกรณีนี้พิจารณาความสัมพันธ์แบบบริบทไม่เชิงเส้น เมื่อพิจารณาจากค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง พบว่าตัวแบบ GCR มีประสิทธิภาพที่ดีที่สุด ตามด้วยตัวแบบ FNN และ CR ตามลำดับ โดยในกรณีนี้เกิดการแก้ไขปัญหาค้นหาเกินพอดีสำหรับตัวแบบ CR อย่างเห็นได้ชัด หากพิจารณาที่จำนวนพารามิเตอร์ที่ใช้ และระยะเวลาในการเรียนรู้ตัวแบบ พบว่าตัวแบบ CR ใช้ทรัพยากรเยอะที่สุด ในขณะที่ตัวแบบ FNN มีจำนวนพารามิเตอร์น้อยกว่าตัวแบบ GCR และเวลาที่ใช้ในการเรียนรู้ต่ำกว่า ซึ่งเป็นผลมาจากความซับซ้อนในโครงสร้างของตัวแบบ ทำให้การคำนวณของตัวแบบ GCR ใช้เวลาที่มากกว่า

จากข้อสรุปข้างต้นทำให้เห็นว่าตัวแบบ FNN เผลอแล้วใช้ทรัพยากรน้อยกว่าตัวแบบ CR แต่ยิ่งมากกว่าตัวแบบ GCR โดยสามารถแก้ไขปัญหาค้นหาเกินพอดีได้บ้าง แต่ไม่ใช่ทุกครั้งเพราะอาจเกิดกรณีที่ค่าถ่วงน้ำหนักเริ่มต้นไม่สามารถทำให้ค่าของฟังก์ชันสูญเสียลู่ออกต่ำสุดได้ ในขณะที่แก้ไขปัญหาค้นหาเกินพอดีไม่ได้เลย

สำหรับตัวแบบ CR เป็นตัวแบบที่ใช้ทรัพยากรเยอะที่สุด จึงสามารถแก้ไขปัญหาค้นหาเกินพอดีได้ดีในระดับหนึ่ง แต่เมื่อมีความซับซ้อนมากขึ้น อาจทำให้เกิดการแก้ไขปัญหาค้นหาเกินพอดีขึ้น เนื่องมาจากมีจำนวนพารามิเตอร์ที่เยอะเกินพอดี

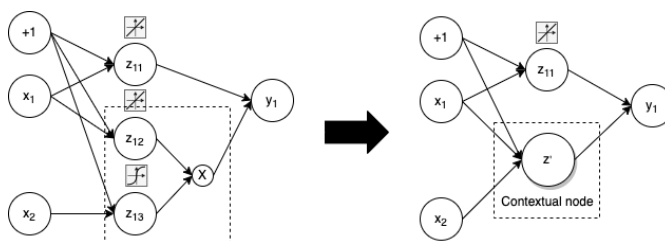
สำหรับตัวแบบ GCR ที่ทางผู้วิจัยได้พัฒนาขึ้นมา นั้น จากผลการทดลอง และข้อสรุป เห็นได้ชัดว่าสามารถแก้ไขปัญหาค้นหาเกินพอดีที่เกิดจากการจำลองข้อมูลแต่ละรูปแบบได้ดีที่สุด โดยให้ค่าเฉลี่ยของค่าเฉลี่ยความผิดพลาดกำลังสอง และใช้เวลาในการเรียนรู้ น้อยที่สุดไม่ว่าจะเป็นปัญหาความสัมพันธ์แบบบริบท

เชิงเส้น หรือไม่เชิงเส้นก็ตาม

5.2 ข้อเสนอแนะ

เนื่องจากงานวิจัยฉบับนี้มีวัตถุประสงค์เพียงเพื่อต้องการประยุกต์ใช้ตัวแบบ GCR ในการแก้ไข ปัญหาการถดถอยเชิงเส้น และเปรียบเทียบกับวิธี FNN และ CR ผ่านการจำลองข้อมูลเท่านั้น โดยทางผู้วิจัยมีข้อเสนอแนะสำหรับผู้อ่านที่สนใจงานวิจัยเพิ่มเติมดังนี้

1. ในงานวิจัยฉบับนี้ มีการกำหนดเงื่อนไขที่ทำการศึกษาไว้ โดยไม่ได้เปลี่ยนเป็นค่าอื่น ซึ่งในแต่ละตัวแบบ แต่ละรูปแบบการจำลองข้อมูล อาจมีการปรับปรุงพารามิเตอร์ที่ใช้ในแต่ละตัวแบบได้ โดย
 - a) ทดลองเปลี่ยนอัตราการเรียนรู้จาก 0.01 เป็นค่าอื่น ๆ
 - b) ทดลองเปลี่ยนจาก Adam optimizer เป็น SGD optimizer หรือ RMSprop optimizer หรือ optimizer อื่น ๆ
 - c) ทดลองเพิ่มจำนวนรอบในการทำ backpropagation ต่อ 1 ตัวแบบ
 - d) ทดลองเปลี่ยนฟังก์ชันกระตุ้นสำหรับตัวแบบ FNN และ CR จากฟังก์ชันไฮเพอร์โบลิกแทน เป็นฟังก์ชันกระตุ้นชนิดอื่นที่ทำให้ลดเวลาเรียนรู้ของตัวแบบได้เช่นฟังก์ชัน ReLu
2. เนื่องจากในงานวิจัยฉบับนี้ ทำการวิเคราะห์ตัวแบบ FNN, CR และ GCR กับข้อมูลที่ได้จำลองขึ้นมาเอง จึงอาจนำไปประยุกต์ใช้กับข้อมูลที่เกิดขึ้นจริงเพื่อวัดประสิทธิภาพตัวแบบเพิ่มเติม
3. สามารถนำวิธีที่ใช้ในตัวแบบ GCR ไปปรับเพิ่มเติมให้เป็น contextual node ได้ด้วยแนวคิดดังรูป 5.1 ซึ่งจะทำหน้าที่เป็นโหนดที่อธิบายความสัมพันธ์ตามบริบทได้



รูปที่ 5.1: แนวคิดในการรวบโหนดที่เกี่ยวข้องกับการวิเคราะห์ตามบริบทให้กลายเป็นโหนดเดียว

ประวัติผู้เขียนวิทยานิพนธ์

ชื่อ-สกุล	นายชยานนท์ ชัตติยาภิรักษ์
วัน เดือน ปี เกิด	2 กันยายน 2538
สถานที่เกิด	กรุงเทพฯ
วุฒิการศึกษา	วิศวกรรมศาสตรบัณฑิต
ที่อยู่ปัจจุบัน	99 ซอยประชาสงเคราะห์ 14 แยก 6 ถนนประชาสงเคราะห์ แขวงดินแดง เขตดินแดง กรุงเทพฯ 10400



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

รายการอ้างอิง

- [1] Noor Ezan Abdullah, Athirah A Rahim, Hadzli Hashim, and Mahanijah Md Kamal, *Classification of rubber tree leaf diseases using multilayer perceptron neural network*, 2007 5th student conference on research and development, IEEE, 2007, pp. 1–6.
- [2] Sebastian Bock, Josef Goppold, and Martin Weiß, *An improvement of the convergence proof of the adam-optimizer*, arXiv preprint arXiv:1804.10587 (2018).
- [3] Josée Dupuis, Josée Coutu, and Odette Laneuville, *Application of linear mixed-effect models for the analysis of exam scores: Online video associated with higher scores for undergraduate students with lower grades*, *Computers & Education* **66** (2013), 64–73.
- [4] Laura M Grajeda, Andrada Ivanescu, Mayuko Saito, Ciprian Crainiceanu, Devan Jagannath, Robert H Gilman, Jean E Crabtree, Dermott Kelleher, Lilia Cabrera, Vitaliano Cama, et al., *Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines*, *Emerging themes in Epidemiology* **13** (2016), no. 1, 1–13.
- [5] Chengyu Liu and Wei Wang, *Contextual regression: an accurate and conveniently interpretable nonlinear model for mining discovery from scientific data*, arXiv preprint arXiv:1710.10728 (2017).
- [6] Jonathan Parkinson, Ryan Hard, Richard I Ainsworth, Nan Li, and Wei Wang, *Engineering a histone reader protein by combining directed evolution, sequencing, and neural network based ordinal regression*, *Journal of Chemical Information and Modeling* **60** (2020), no. 8, 3992–4004.
- [7] Sagar Sharma, Simone Sharma, and Anidhya Athaiya, *Activation functions in neural networks, towards data science* **6** (2017), no. 12, 310–316.
- [8] Michael W Sjoding, Robert P Dickson, Theodore J Iwashyna, Steven E Gay, and Thomas S Valley, *Racial bias in pulse oximetry measurement*, *New England*

Journal of Medicine **383** (2020), no. 25, 2477–2478.

- [9] Fabrizzio Alphonsus AMN Soares, Edna Lúcia Flôres, Christian Dias Cabacinha, Gilberto Arantes Carrijo, and Antônio Cláudio Paschoarelli Veiga, *Recursive diameter prediction and volume calculation of eucalyptus trees using multilayer perceptron networks*, Computers and electronics in agriculture **78** (2011), no. 1, 19–27.
- [10] M Talaat, MA Farahat, Noura Mansour, and AY Hatata, *Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach*, Energy **196** (2020), 117087.

