



1.1 ความเป็นมาและความสำคัญของปัญหา

งานประเภทปัญญาประดิษฐ์ (Artificial Intelligence) และการเรียนรู้ของเครื่อง (Machine Learning) นั้นเป็นงานที่เกี่ยวข้องกับการพยายามที่จะทำให้คอมพิวเตอร์สามารถทำการจดจำหรือเรียนรู้ความรู้บางอย่างขึ้นมาได้เหมือนกับมนุษย์ซึ่งเป็นสิ่งมีชีวิตที่มีภูมิปัญญา ยกตัวอย่างเช่น การที่มนุษย์สามารถมองเห็นสิ่งต่างๆ แล้วเรียนรู้คุณสมบัติของสิ่งเหล่านั้น หรือรู้สึกได้ถึงความคล้ายคลึงกันระหว่างสิ่งของ อย่างไรก็ตาม แม้แต่กลไกอันซับซ้อนต่างๆ ที่ถือได้ว่าเป็นลักษณะการเรียนรู้ที่แสดงถึงสติปัญญาเหล่านี้ก็ย่อมที่จะสามารถอธิบายได้โดยอาศัยกฎเกณฑ์หรืออัลกอริทึมบางอย่าง หากเราทราบถึงกฎเกณฑ์ที่น่าจะดีที่สุดสำหรับงานเหล่านี้หรือกฎเกณฑ์ที่มนุษย์ใช้ในการตัดสินใจอย่างแน่ชัดแล้ว เราก็ย่อมสามารถทำให้คอมพิวเตอร์ซึ่งเป็นเครื่องจักรที่เราประดิษฐ์ขึ้นทำงานตามกฎเกณฑ์เหล่านั้นได้ เช่นนี้เอง คอมพิวเตอร์ก็สามารถที่จะแสดงลักษณะของการเรียนรู้ที่สะท้อนถึงสติปัญญาได้เช่นกันแม้ว่ามันจะเป็นเพียงวัตถุที่ไร้ซึ่งจิตวิญญาณและความคิด ด้วยเหตุนี้ เราจึงเรียกสิ่งที่เกิดขึ้นว่า ปัญญาประดิษฐ์ นั่นก็คือเราสามารถสร้างสิ่งที่เป็นเหมือนสติปัญญาเทียม ขึ้นมาได้ และด้วยการผสมเข้าด้วยกันระหว่างกฎเกณฑ์ที่ใช้ในการตัดสินใจของมนุษย์กับการทำงานได้โดยไม่รู้จักเหน็ดเหนื่อยและแทบจะไร้ซึ่งข้อผิดพลาดของคอมพิวเตอร์ เราจึงสามารถที่จะทำการประมวลผลโดยใช้สติปัญญาได้กับงานที่มีข้อมูลจำนวนมาก หรือเป็นงานที่ต้องอาศัยความพยายามเกินกว่าระดับที่มนุษย์ซึ่งเป็นสิ่งมีชีวิตจะสามารถกระทำได้โดยอาศัยสติปัญญาเทียมนี้

ในงานจำพวกปัญญาประดิษฐ์และการเรียนรู้ของเครื่องนั้นสามารถแยกย่อยออกไปได้เป็นหลายอย่าง และหนึ่งในงานเหล่านั้นก็คืองานที่เรียกว่าการจำแนกประเภท การจำแนกประเภทที่กล่าวถึงนี้ก็คือการหาวิธีการเพื่อที่จะนำมาใช้ในการทำนายประเภท (class) ของข้อมูลที่สนใจโดยอาศัยคุณสมบัติ (attribute) ของข้อมูลที่มีอยู่ในเบื้องต้น ยกตัวอย่างเช่น ในการที่จะทำนายเพศของบุคคลโดยอาศัยคุณสมบัติของข้อมูลเกี่ยวกับส่วนสูง น้ำหนัก และอายุ เริ่มต้นเราจะมีข้อมูลจำนวนหนึ่งเกี่ยวกับ เพศ ส่วนสูง น้ำหนักและอายุของบุคคลต่างๆ เมื่อทำการใช้อัลกอริทึมในการจำแนกประเภทกับข้อมูลเบื้องต้นเหล่านั้นแล้วอัลกอริทึมนี้จะให้ผลลัพธ์ออกมาเป็นวิธีการในการจำแนกประเภทหรือทำนายประเภทของข้อมูลที่สนใจ โดยในที่นี้จะเป็วิธีการที่จะทำนายเพศโดยอาศัยคุณสมบัติของข้อมูลเกี่ยวกับส่วนสูง น้ำหนัก และ อายุ

มีอัลกอริทึมอยู่มากมายหลายอย่างที่ถูกรู้จักกันขึ้นมาสำหรับงานของการจำแนกประเภทนี้ และหนึ่งในอัลกอริทึมที่รู้จักกันดีซึ่งถือได้ว่าเป็นอัลกอริทึมพื้นฐานนั่นก็คืออัลกอริทึมเน็ยเรสท์เนเบอร์ (nearest neighbor algorithm) [1,2] อัลกอริทึมเน็ยเรสท์เนเบอร์เป็นอัลกอริทึมสำหรับการ

จำแนกประเภทซึ่งตั้งอยู่บนพื้นฐานที่จะทำการจำแนกประเภทข้อมูลโดยอาศัยค่าของประเภทตามข้อมูลตัวอย่างที่มีลักษณะคล้ายคลึงกัน อัลกอริทึมเนียบเรสท์เนเบอร์แบบดั้งเดิมหรือที่รู้จักกันในนามเคเนียบเรสท์เนเบอร์นั้นมีพารามิเตอร์ 1 ตัวคือ k ซึ่งจะเป็นจำนวนเต็มบวกที่ใช้กำหนดจำนวนข้อมูลตัวอย่างกลุ่มหนึ่งที่มีความคล้ายคลึงกับข้อมูลที่สนใจที่ต้องการทำนายประเภทมากที่สุดและจะถูกนำมาใช้ในการจำแนกประเภทของข้อมูลที่สนใจ โดยในท้ายที่สุดการทำนายผลก็จะอาศัยการเลือกประเภทที่ปรากฏอยู่มากที่สุดจากข้อมูลคล้ายคลึงเหล่านั้น นั่นก็คือการใช้เสียงส่วนใหญ่ (majority vote)

อัลกอริทึมเนียบเรสท์เนเบอร์ถ่วงน้ำหนัก (weighted nearest neighbor) เป็นอัลกอริทึมแบบหนึ่งที่ยกย่อยออกมาจากอัลกอริทึมเนียบเรสท์เนเบอร์ โดยดัดแปลงจากเคเนียบเรสท์เนเบอร์โดยการใช้ค่าถ่วงน้ำหนักให้กับข้อมูลต่างๆ อัลกอริทึมนี้จะอาศัยพื้นฐานที่ว่าข้อมูลที่มีความคล้ายคลึงกันมากกว่าควรจะมีค่าสำคัญหรือส่งผลต่อผลลัพธ์ในการทำนายประเภทมากกว่าข้อมูลอื่น อัลกอริทึมนี้จะทำตามพื้นฐานดังกล่าว นั่นก็คือการใช้ฟังก์ชันที่รับอินพุตเป็นความแตกต่างระหว่างข้อมูลทั้งสองแล้วให้ผลลัพธ์ออกมาเป็นค่าซึ่งแทนความสำคัญของข้อมูลเหล่านั้นต่อการจำแนกประเภทข้อมูลที่สนใจ ในการคำนวณความแตกต่างระหว่างข้อมูลที่ใช้เป็นอินพุตของฟังก์ชันถ่วงน้ำหนักนั้นสามารถคำนวณได้โดยการถือว่าคุณสมบัติต่างๆ ของข้อมูลเป็นมิติต่างๆ แล้วคำนวณหาระยะทางระหว่างจุด 2 จุด เมื่อพิจารณาว่าจุดเหล่านั้นเป็นตัวแทนข้อมูลในปริภูมิซึ่งมีจำนวนมิติเท่ากับจำนวนคุณสมบัติของข้อมูลนั้น โดยตัวอย่างของฟังก์ชันถ่วงน้ำหนักที่เป็นที่รู้จักกันดีก็คือ $f(x) = \frac{1}{x^2}$ เมื่อ x เป็นระยะทางระหว่างจุด 2 จุดซึ่งแสดงถึงความแตกต่างของข้อมูลทั้งสอง [1,2]

งานวิจัยนี้เกี่ยวข้องกับอัลกอริทึมเนียบเรสท์เนเบอร์ที่ได้กล่าวถึงมาแล้ว โดยงานวิจัยนี้จะพยายามค้นหาฟังก์ชันถ่วงน้ำหนักซึ่งมีคุณสมบัติที่คิดตามที่ต้องการและเพื่อยืนยันประสิทธิภาพของฟังก์ชันถ่วงน้ำหนักต่างๆ ก็จะอาศัยการนำฟังก์ชันที่ได้ไปทดสอบประสิทธิภาพ โดยในการวัดผลนั้นจะอาศัยการเปรียบเทียบกันระหว่างความแม่นยำเมื่อใช้ฟังก์ชันที่นำเสนอเปรียบเทียบกับความแม่นยำที่ได้จากฟังก์ชันเดิม

1.2 วัตถุประสงค์ของการวิจัย

งานวิจัยนี้มีวัตถุประสงค์หลักเพื่อที่จะหาฟังก์ชันถ่วงน้ำหนักสำหรับนำมาใช้อัลกอริทึมเนียบเรสท์เนเบอร์แบบถ่วงน้ำหนักซึ่งคาดว่าจะมีคุณสมบัติเหมาะสมและให้ความถูกต้องในการจำแนกประเภทหรือที่เรียกว่าความแม่นยำมากกว่าผลที่ได้จากฟังก์ชันถ่วงน้ำหนักมาตรฐานทั่วไป

1.3 ขอบเขตของวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้สนใจที่จะพัฒนาประสิทธิภาพอัลกอริทึมเนียร์เรสต์เนเบอร์แบบถ่วงน้ำหนักสำหรับงานในการจำแนกประเภทโดยไม่สนใจการประมาณค่าฟังก์ชันที่ต้องการผลลัพธ์เป็นค่าต่อเนื่อง โดยเน้นเฉพาะในส่วนของฟังก์ชันถ่วงน้ำหนักที่จะถูกนำไปใช้กับอัลกอริทึมโดยที่ไม่เกี่ยวข้องกับเทคนิคที่จะพัฒนาความแม่นยำในด้านอื่น และในการวัดหรือทดสอบประสิทธิภาพก็จะเน้นการเปรียบเทียบความแม่นยำที่ได้เป็นหลักโดยไม่สนใจปัจจัยอื่นๆ เช่น ความรวดเร็วในการทำงาน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

วิทยานิพนธ์นี้จะให้ผลลัพธ์เป็นฟังก์ชันถ่วงน้ำหนักสำหรับอัลกอริทึมเนียร์เรสต์เนเบอร์ซึ่งมีคุณสมบัติต่างๆ ที่มีความเหมาะสมหรือดีทั้งในทางทฤษฎีและทางปฏิบัติ นั่นก็คือน่าจะได้รับฟังก์ชันที่สามารถพิสูจน์ได้โดยทางทฤษฎีว่ามีประสิทธิภาพดีและเมื่อนำไปทำการทดสอบโดยการทดลองใช้ในการจำแนกประเภทของชุดข้อมูลต่างๆ ซึ่งถือว่าเป็นทางปฏิบัติ ก็จะทำให้ผลลัพธ์คือความแม่นยำที่ดีเหนือกว่าตัวเปรียบเทียบ

1.5 วิธีดำเนินการวิจัย

วิธีดำเนินการวิจัยสามารถสรุปเป็นข้อๆ ได้ดังต่อไปนี้

1. ศึกษาความรู้ที่เกี่ยวข้อง ซึ่งในที่นี้ก็คือความรู้ที่เกี่ยวข้องกับอัลกอริทึมเนียร์เรสต์เนเบอร์และฟังก์ชันถ่วงน้ำหนักต่างๆ
2. คิดค้นรูปแบบของฟังก์ชันถ่วงน้ำหนักที่ดี ซึ่งก็คือการทำงานในทางทฤษฎี
3. ทำการทดลองเพื่อทดสอบประสิทธิภาพของฟังก์ชันถ่วงน้ำหนักที่ได้ โดยการเขียนโปรแกรมสำหรับทดสอบประสิทธิภาพของฟังก์ชันที่คิดค้นขึ้นมาในการทำงานจริง
4. สรุปผลการทดลองทั้งหมดและจัดทำวิทยานิพนธ์

