

บทที่ 3

โครงข่ายประสาทเทียม

ในปัจจุบันนี้คอมพิวเตอร์ได้ถูกนำมาใช้งานอย่างกว้างขวางเกือบจะทุกด้านก็ว่าได้ แต่ยังคงมีขีดความสามารถจำกัดอยู่เมื่อเทียบกับสมองมนุษย์ ซึ่งสมองสามารถจดจำและเรียนรู้จากประสบการณ์ในอดีตและนำไปปรับใช้กับสถานการณ์ปัจจุบัน ดังนั้นจึงได้มีการศึกษาการทำงานของสมอง เพื่อนำมาเป็นแบบจำลองของเซลล์ประสาท (Artificial Neural Model) โครงข่ายประสาทเทียม (Artificial Neural Network) และศึกษากระบวนการเรียนรู้ (Learning Algorithm) เพื่อนำมาประยุกต์ใช้กับเครื่องคอมพิวเตอร์ ซึ่งการนำทฤษฎีต่างๆ ของโครงข่ายประสาทที่ผ่านการพัฒนามาหลายสิบปีมาประยุกต์ใช้กับงานจริงมีอย่างมากมายหลายด้าน เช่น โทรคมนาคม นำมาใช้ในการบีบข้อมูลและภาพ (Image and Data Compression) ระบบนักบินอัตโนมัติ (Auto Pilot Aircraft) การแปลภาษาพูด (Real Time Translation of spoken language) การจดจำเสียง (Voice recognition) การจดจำภาพ (Image recognition) การจดจำรูปแบบ (Pattern recognition) ซึ่งจะจดจำแบบเดิมและจำรูปแบบนั้นเมื่อพบอีกครั้ง นอกจากนี้ยังมีการใช้งานแบบอื่นๆ เช่น การค้นหาเป้าหมาย (Trajectory Control) การวิเคราะห์การตลาด (Market Analysis) ฯลฯ นักวิจัยด้านไฟฟ้ากำลังจึงมีแนวคิดที่ประยุกต์ใช้โครงข่ายประสาทเทียมเข้ามาแก้ปัญหาเกี่ยวกับไฟฟ้ากำลังหลายปัญหาด้วยกัน เช่น ปัญหาโหลดฟลว (Load Flow) , การพยากรณ์ความต้องการกำลังไฟฟ้า (Load Forecasting) และปัญหาการจ่ายโหลดอย่างมีประสิทธิภาพ (Economic Load Dispatch) เป็นต้น ในบทนี้จะเป็นการอธิบายหลักการของโครงข่ายประสาทเทียม โดยจะกล่าวถึงโครงข่ายประสาทเทียม, ชนิดของการเรียนรู้ของโครงข่ายประสาทเทียม รวมถึงโครงข่ายประสาทเทียมชนิดต่างๆ เพื่อนำมาประยุกต์ใช้ในงานวิจัย

3.1 ความเป็นมาของโครงข่ายประสาทเทียม [8]

ความพยายามที่จะทำความเข้าใจการทำงานของสมองมนุษย์ได้มีมานานนับศตวรรษแล้ว เริ่มต้นในปี ค.ศ. 1890 William James นักสรีรศาสตร์ชาวอเมริกันได้แต่งหนังสือชื่อ Psychology (Brief Course) ซึ่งอธิบายถึงการทำงานของสมอง และยังกล่าวถึงทฤษฎีที่มีอยู่ในปัจจุบันไว้ล่วงหน้าปี ค.ศ. 1936 Alan Turing ได้นำสมองไปใช้เป็นแบบอย่างของการคำนวณใน A way of looking at the world of computing ในปี ค.ศ. 1943 Warren McCulloch นักประสาทวิทยาและ Walter Pitts นักคณิตศาสตร์ซึ่งมีอายุเพียง 18 ปี ได้ออกแบบและเป็นที่ให้ความสนใจโดยทั่วไปว่าเป็นโครงข่ายประสาทเทียมชนิดแรก นักวิจัยเหล่านั้นยอมรับว่าการรวมหลายๆ นิวรอนไม่ซับซ้อนเข้าสู่ระบบโครงข่ายให้เป็นการเพิ่มกำลังในการคำนวณค่าถ่วงน้ำหนักของนิวรอน โดยมีแนวคิด

สำคัญคือการคำนวณค่าสัญญาณเอาต์พุตโดยพิจารณาจากระดับกระตุ้นกล่าวคือ นิวรอนจะให้สัญญาณเอาต์พุตก็ต่อเมื่อ ผลรวมของสัญญาณอินพุตมีค่ามากกว่าระดับกระตุ้นซึ่งแนวคิดนี้ยังคงเป็นส่วนสำคัญในการพัฒนาโครงข่ายประสาทเทียมมาจนทุกวันนี้

ในปี 1949 Dinal Hebb นักฟิสิกส์จากคณะฟิสิกส์มหาวิทยาลัย McGill ได้สร้างกฎการเรียนรู้ของโครงข่ายประสาทเทียมขึ้นเป็นครั้งแรก ทฤษฎีของ Hebb กล่าวว่าถ้านิวรอน 2 หน่วยทำงานพร้อมกันค่าถ่วงน้ำหนักระหว่างสายการเชื่อมโยงของทั้งสองนิวรอนต้องลดลง

ในช่วงปี ค.ศ.1950 ถึง ค.ศ.1960 ถือว่าเป็นช่วงที่มีการพัฒนาโครงข่ายประสาทเทียมกันอย่างกว้างขวาง เมื่อ John van Neumann บิดาแห่งวงการคอมพิวเตอร์ยุคใหม่ได้นำเอาแบบจำลองของสมองที่เสนอโดยนักวิจัยที่ทำงานเกี่ยวกับโครงข่ายประสาทเทียมไม่ว่าจะเป็น Warren McCulloch และคนอื่น ๆ มาเป็นแนวทางในการพัฒนาเครื่องคอมพิวเตอร์ในยุคต่อมา จนในปี ค.ศ.1956 ก็ได้มีการสร้างแบบจำลองโดยใช้คอมพิวเตอร์ขึ้นได้สำเร็จ

ในปี ค.ศ.1957 Frank Rosenblatt ได้เริ่มศึกษาและพัฒนาประเภทใหญ่ของโครงข่ายประสาทเทียมที่เรียกว่า Perceptron ซึ่งเป็นรูปแบบที่สำคัญรูปแบบหนึ่งของโครงข่ายประสาทเทียม โดยส่วนมากสถาปัตยกรรมของโครงข่ายประสาทเทียมชนิดนี้จะประกอบด้วยชั้นอินพุตเชื่อมโยงผ่านด้วยค่าถ่วงน้ำหนัก เพื่อเป็นตัวเชื่อมโยงกับนิวรอนที่ชั้นเอาต์พุต ค่าถ่วงน้ำหนักบนทางผ่านเชื่อมโยงถูกปรับเปลี่ยน กฎการเรียนรู้ Perceptron ปรับเปลี่ยนค่าถ่วงน้ำหนักซ้ำ ๆ กันซึ่งมีประสิทธิภาพดีกว่ากฎของ Hebb การเรียนรู้ของ Perceptron สามารถแสดงให้เห็นว่าลู่เข้าสู่ค่าถ่วงน้ำหนักที่ถูกต้อง ถ้าค่าถ่วงน้ำหนักเหล่านั้นจะแก้ปัญหาก็เตรียมพร้อม

ในปี ค.ศ.1959 Bernard Widrow และนักศึกษาของเขา Marcian Hoff ได้พัฒนาการเรียนรู้ซึ่งใกล้เคียงกับกฎการเรียนรู้ของ Perceptron โดยกฎการเรียนรู้ของ Perceptron ปรับค่าถ่วงน้ำหนักที่เชื่อมโยงที่โหนดเมื่อผลตอบสนองของโหนดไม่ถูกต้อง กฎ Delta ปรับค่าถ่วงน้ำหนักเพื่อลดความแตกต่างระหว่างโครงข่ายอินพุตถึงโหนดเอาต์พุตและค่าเอาต์พุตที่ต้องการ สิ่งเหล่านี้จะสรุปในรูปแบบของค่าผิดพลาดยกกำลังสองเฉลี่ยที่น้อยที่สุด (Smallest Mean Square Error : SMSE) และนำสิ่งต่าง ๆ เหล่านี้ไปใช้สร้างแบบจำลองของ MADALINE (Multiple ADaptive LINear Element) ซึ่งพัฒนามาจาก ADALINE และเป็นการนำเอาโครงข่ายประสาทเทียมมาใช้งานในทางปฏิบัติเป็นครั้งแรก โดยนำมาลดเสียงก้องในสายโทรศัพท์

ในปี ค.ศ. 1969 Marvin Minsky และ Seymour Papert ได้ตีพิมพ์หนังสือ Perceptron ซึ่งกล่าวถึงข้อจำกัดของ Perceptron ว่าไม่สามารถนำไปใช้ในการแก้ปัญหามากมาย ปัญหาที่มีความน่าสนใจได้ ซึ่งจุดนี้เองทำให้นักวิจัยหลายคนกลับไปให้ความสนใจในด้านปัญญาประดิษฐ์ (AI) มากขึ้น แต่ยังมีนักวิจัยบางกลุ่มที่ไม่ย่อท้อและยังคงทำงานวิจัยด้านโครงข่ายประสาทเทียมต่อไป

ในปี ค.ศ.1968 John Anderson ของมหาวิทยาลัย Brown เริ่มวิจัยงานของเขาในโครงข่ายประสาทเทียมด้วยโครงข่ายรวมกลุ่มความสามารถในการจดจำ (Associative Memory Nets) เขาได้

พัฒนาแบบจำลองของโครงข่ายประสาทเทียมขึ้นมา และให้ชื่อว่า Brain-State-in-a-Box (BSB) งานประยุกต์ที่ใช้ในโครงข่ายประสาทเทียมชนิดนี้คือ การวินิจฉัยทางการแพทย์ (Medical Diagnosis)

ในปี ค.ศ.1972 Teuvo Kohonen วิศวกรไฟฟ้าชาวฟินแลนด์ของมหาวิทยาลัยเทคโนโลยี Helsinki ได้พัฒนาการเรียนรู้ของโครงข่ายประสาทเทียมด้วยการแก้ไขค่าถ่วงน้ำหนักแบบเมตริกซ์ โดยอาศัยพื้นฐานจากทฤษฎีของ Hebb และในปีเดียวกันเขาก็ได้เสนอทฤษฎีการเรียนรู้ของโครงข่ายประสาทเทียมแบบ Self - Organizing map โดยการจัดรวมนิวรอนเป็นกลุ่มเพื่อใช้ในการวิเคราะห์เสียงพูด

ในช่วงปี ค.ศ.1985 ถึง ค.ศ.1990 Gail Carpenter ได้ร่วมวิจัยกับ Stephen Grossberg ได้เสนอทฤษฎี Adaptive Resonance สำหรับวิเคราะห์ข้อมูลแบบไบนารี (ART1) และข้อมูลอินพุตแบบต่อเนื่อง (ART2) และในปี ค.ศ. 1985 David Parker ได้เสนอทฤษฎีการเรียนรู้แบบแพร่ค่าย้อนกลับ (Back-propagation) ทฤษฎีการเรียนรู้นี้ใช้การนำค่าผิดพลาดของสัญญาณเอาต์พุตที่คำนวณได้กับสัญญาณเอาต์พุตตัวอย่างกลับมาเป็นค่าแก้ไขของค่าถ่วงน้ำหนัก

ในปัจจุบันโครงข่ายประสาทเทียมได้ถูกพัฒนา ค้นคว้า วิจัยและนำไปใช้ในด้านต่าง ๆ มากขึ้น อาทิ วิทยาศาสตร์ด้านระบบประสาท, การเงินการธนาคาร, ด้านคณิตศาสตร์, การแพทย์, ระบบสื่อสาร รวมไปถึงงานทางด้านวิศวกรรมไฟฟ้าอีกด้วย จึงถือได้ว่าโครงข่ายประสาทเทียมเป็นศาสตร์แขนงหนึ่งที่มีความสนใจเป็นอย่างมากในช่วงเวลานี้

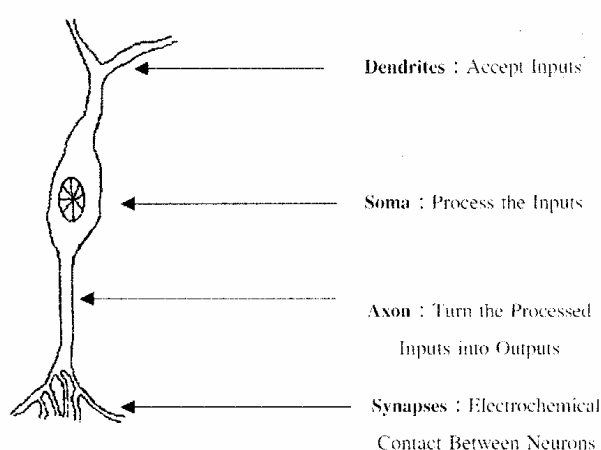
3.2 โครงข่ายประสาทชีวภาพ [9]

การทำงานที่แท้จริงของสมองมนุษย์ยังคงเป็นเรื่องลึกลับ แม้กระนั้นบางแง่มุมของตัวประมวลผลที่มหัศจรรย์นี้เป็นที่ทราบกัน องค์ประกอบพื้นฐานที่สุดของสมองมนุษย์ คือ ประสาทเฉพาะของเซลล์ที่ไม่เหมือนกันกับเซลล์ที่เหลือของทั้งร่างกายซึ่งมันจะไม่สร้างขึ้นใหม่อีก เพราะเซลล์ประสาทนี้เป็นส่วนเดียวของร่างกายที่ไม่ถูกแทนที่ มันถูกสันนิษฐานว่าเซลล์เหล่านี้คือ สิ่งที่ทำให้เราสามารถที่จะจดจำคิดและประยุกต์ใช้ประสบการณ์จากก่อนหน้ากับทุก ๆ การกระทำของเรา เซลล์เหล่านี้ (ทั้งหมด 100 พันล้านเซลล์) เป็นที่รู้จักในฐานะเซลล์ประสาทหรือนิวรอน (Neuron) แต่ละนิวรอนจะสามารถเชื่อมต่อกันกับนิวรอนอื่น ๆ ได้มากถึง 200,000 นิวรอน โดยปกติแล้วจะเป็น 1,000 ถึง 10,000 เซลล์ พลังอำนาจของสติปัญญามนุษย์ มาจากจำนวนเล็กน้อยขององค์ประกอบพื้นฐานเหล่านี้ จากการเชื่อมต่อระหว่างกันและการเรียนรู้ซึ่งเกี่ยวข้องกับพันธกรรมด้วย นิวรอนมีความซับซ้อน มันมีส่วนประกอบระบบย่อยและกลไกควบคุมจำนวนมาก นิวรอนมีกลุ่มที่แตกต่างกันมากกว่า 100 กลุ่มขึ้นอยู่กับวิธีการที่ใช้จำแนกเมื่อร่วมกัน นิวรอนเหล่านี้และการเชื่อมต่อของมันสร้างกระบวนการที่ไม่ใช่ไบนารี ไม่สม่ำเสมอและไม่เป็นจังหวะเดียวกันกล่าวคือมันไม่เหมือนกับคอมพิวเตอร์ที่มีในปัจจุบันหรือแม้แต่โครงข่ายประสาทเทียม

3.2.1 โครงสร้างของนิวรอน

นิวรอนแบ่งออกเป็น 4 ส่วน ดังรูปที่ 3.1 แต่ละส่วนมีฟังก์ชันการทำงานดังนี้

1. เดนไดรต์ (Dendrite) คือ ส่วนที่ทำหน้าที่รับข้อมูลเข้า
2. โซมา (Soma) คือ ส่วนที่ประมวลผลข้อมูลเบื้องต้น
3. แอกซอน (Axon) คือ ส่วนที่แปลงข้อมูลที่ได้ประมวลผลเบื้องต้นเป็นผลลัพธ์ที่ต้องการ
4. ซิแนปส์ (Synapse) คือ เส้นประสาทที่ทำหน้าที่เชื่อมต่อเพื่อการสื่อสารกับนิวรอนอื่น



รูปที่ 3.1 แสดงโครงสร้างของนิวรอน [9]

3.2.2 การส่งผ่านกระแสประสาท [8]

กระแสประสาทจะถูกส่งผ่านจากนิวรอนหนึ่งไปยังอีกนิวรอนหนึ่งโดยผ่านไซแนปส์ เมื่อกระแสประสาทผ่านเข้าไปในไซแนปส์ สารเคมีในไซแนปส์จะมีผลต่อกระแสประสาทได้สองลักษณะ คือการกระตุ้นให้กระแสประสาทมีขนาดเพิ่มขึ้น หรือยับยั้งให้กระแสประสาทมีขนาดลดลง ซึ่งทั้งสองลักษณะดังกล่าวเปรียบได้กับการที่สัญญาณอินพุตถูกถ่วงน้ำหนักให้มีค่ามากขึ้นหรือลดลง เมื่อผ่านออกจากไซแนปส์กระแสประสาทที่ถูกกระตุ้นและกระแสประสาทที่ถูกยับยั้งจะถูกรวมเข้าด้วยกัน หากผลรวมของกระแสประสาทดังกล่าวมีค่ามากกว่าหรือเท่ากับศักยค่าที่ผิวของนิวรอนถัดไป หรือเทรชโฮลด์ (Threshold) กระแสประสาทดังกล่าวก็จะสามารถไหลผ่านเข้าไปในนิวรอนได้ หากผลรวมของกระแสประสาทนั้นมีค่าน้อยกว่าค่าเทรชโฮลด์ กระแสประสาทก็จะถูกยับยั้งไม่ให้ผ่านเข้าไปในนิวรอนนั้น การไหลของกระแสประสาทจะเป็นไปในทิศทางเดียว และมีความเร็วอยู่ระหว่าง 10 ถึง 120 เมตรต่อวินาที เมื่อการส่งกระแสประสาทครั้งหนึ่งสิ้นสุดลง ส่วนต่างๆของนิวรอนจะกลับเข้าสู่สภาวะเริ่มต้นอีกครั้งหนึ่งเพื่อเตรียมพร้อมสำหรับส่งกระแสประสาทครั้งต่อไป อาจสรุปคุณสมบัติของโครงข่ายประสาทชีวภาพได้ดังนี้

- 1) แต่ละนิวรอนจะรับข้อมูลอินพุตเป็นจำนวนมาก
- 2) ข้อมูลอินพุตอาจถูกปรับให้เหมาะสมด้วยค่าการเชื่อมโยง (Connection Value of Weight Value of Strength) ที่ไซแนปส์
- 3) ข้อมูลอินพุตที่ผ่านการปรับด้วยค่าการเชื่อมโยงจะถูกรวมและทำการประมวลผล
- 4) ภายใต้ภาวะที่เหมาะสม เช่น ข้อมูลอินพุตจำนวนเพียงพอ นิวรอนจะส่งข้อมูลออกมาหนึ่งค่า
- 5) ข้อมูลเอาต์พุตจากนิวรอนหนึ่งอาจจะถูกส่งไปที่นิวรอนอื่นๆหลายนิวรอนผ่านทางแอกซอน
- 6) นิวรอนแต่ละนิวรอนมีค่าเป็นเฉพาะตำแหน่งหรือเฉพาะที่ (Local)
- 7) ความจำแบ่งออกเป็น 2 ส่วนคือ
 - ความจำระยะยาว จะเก็บที่ไซแนปส์ของนิวรอนหรือเก็บในค่าการเชื่อมโยง
 - ความจำระยะสั้น สัมพันธ์กับข้อมูลเอาต์พุตจากนิวรอน
- 8) ความถูกต้องของการเชื่อมโยงขึ้นกับประสบการณ์ในการเรียนรู้ที่ผ่านมา
- 9) การควบคุมการถ่ายโอนของไซแนปส์อาจจะเป็นการกระตุ้นแบบเสริมหรือแบบหักล้าง

นอกจากที่กล่าวมา โครงข่ายประสาทชีวภาพยังยืดหยุ่นต่อความผิดพลาด (Fault – Tolerance) 2 ประเด็นด้วยกัน ประการแรกคือ ความสามารถในการจดจำและจำแนกข้อมูลอินพุตที่แตกต่างจากข้อมูลอินพุตที่เคยได้รับไม่มากนัก ตัวอย่างเช่นความสามารถในการจำแนกภาพบุคคลที่ไม่เคยเห็นมาก่อน หรือสามารถจดจำบุคคลที่ไม่ได้พบเห็นกันเป็นเวลานาน ประการที่สองความสามารถในการซ่อมแซมหรือชดเชยความเสียหายของตัวนิวรอนเอง ในสมองมนุษย์เมื่อนิวรอนหนึ่งชำรุดเสียหายก็จะมีสร้างนิวรอนใหม่ขึ้นมาแทนที่ นั่นก็ยังคงมีการเรียนรู้อย่างต่อเนื่องแม้ว่าจะสูญเสียนิวรอนบางส่วน ในบางครั้งนิวรอนที่เหลือจะได้รับการฝึกสอนเพื่อทำหน้าที่แทนนิวรอนที่ชำรุดเสียหาย

จากจุดเด่นหลายประการของโครงข่ายประสาทชีวภาพผนวกกับเทคโนโลยีด้านคอมพิวเตอร์ การทำงานแต่ละครั้งของอุปกรณ์สารกึ่งตัวนำในคอมพิวเตอร์จะใช้เวลาอยู่ในช่วงนาโนวินาที (10^{-9} s) ซึ่งมีความรวดเร็วมากเมื่อเปรียบเทียบกับการทำงานของสมองมนุษย์ทำให้มีการคิดค้นแนวทางแก้ปัญหาโดยมีโครงสร้างองค์ประกอบและกลไกการแก้ปัญหาลักษณะเดียวกับโครงข่ายประสาทชีวภาพ โดยเรียกว่า โครงข่ายประสาทเทียม (Artificial Neural Networks)

3.3 โครงข่ายประสาทเทียม [9]

โครงข่ายประสาทเทียม คือแบบจำลองอิเล็กทรอนิกส์ที่ยังไม่สมบูรณ์ซึ่งมีพื้นฐานบนโครงข่ายธรรมชาติของสมอง ความสัมพันธ์ระหว่างเซลล์ประสาทกับเซลล์ประสาทเทียมเป็นดังนี้

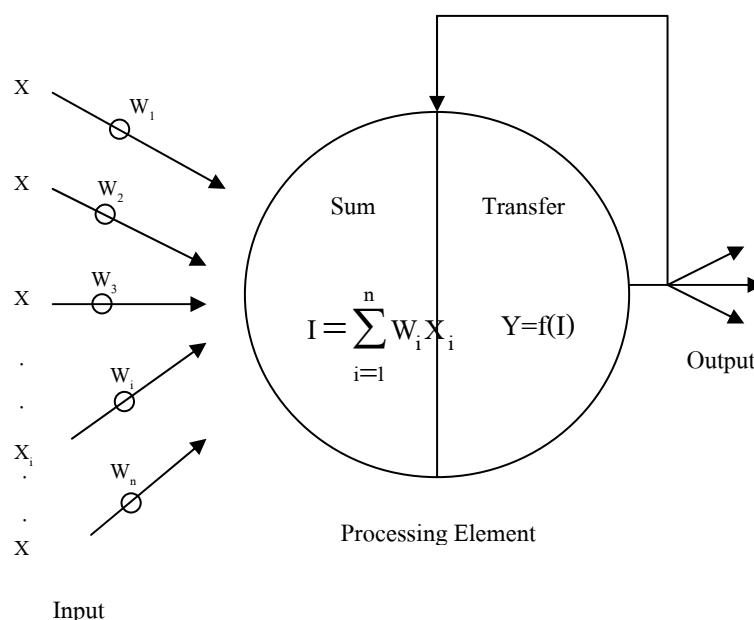
เซลล์ประสาท

- ตัวเซลล์ (Cell Body)
- เดนไดรต์ (Dendrite)
- แอกซอน (Axon)
- ซิแนปส์ (Synapse)

เซลล์ประสาทเทียม

- ยูนิต (Unit)
- ชั้นข้อมูลด้านเข้า (Input Layer)
- ชั้นข้อมูลด้านออก (Output Layer)
- ค่าถ่วงน้ำหนัก (Weight Value)

โครงสร้างการทำงานของนิวรอนในโครงข่ายประสาทเทียมเป็นดังรูปที่ 3.2



รูปที่ 3.2 แสดงโครงสร้างการทำงานของนิวรอนในโครงข่ายประสาทเทียม [9]

จากรูปที่ 3.2 ถ้ากำหนดให้ข้อมูลที่ป้อนเข้าสู่โครงข่ายประสาทเทียม คือ $X_1, X_2, \dots, X_n$ ข้อมูลแต่ละตัวจะถูกคูณด้วยค่าถ่วงน้ำหนัก (Weight) $W_1, W_2, \dots, W_n$ ตามลำดับ จากนั้นข้อมูลแต่ละตัวก็จะถูกรวมเข้าด้วยกันซึ่งกระบวนการนี้เรียกว่า Summation ก่อนที่จะแปลงข้อมูล (Transfer) เป็นผลลัพธ์แล้วส่งให้นิวรอนตัวอื่นนำไปประมวลผลในลักษณะเดียวกันเพื่อสร้างผลลัพธ์ขั้นสุดท้าย

โครงข่ายประสาทเทียมอาจพิจารณาให้เป็นขั้นตอนวิธีทางคณิตศาสตร์ (Mathematical Algorithm) ของโครงข่ายประสาทชีวภาพโดยตั้งบนพื้นฐานดังนี้

1. หน่วยประมวลผลเรียกว่า นิวรอน การส่งผ่านข้อมูลระหว่างนิวรอนทำได้โดยการเชื่อมโยงระหว่างนิวรอน
2. การเชื่อมโยงแต่ละส่วนจะขึ้นอยู่กับค่าการเชื่อมโยงหรือค่าถ่วงน้ำหนัก

ข้อมูลเอาต์พุตจะถูกกำหนดด้วยฟังก์ชันกระตุ้น (Activation function) ซึ่งปกติความสัมพันธ์ระหว่างผลรวมของข้อมูลอินพุตที่คูณด้วยค่าการเชื่อมโยงกับข้อมูลเอาต์พุตเป็นความสัมพันธ์แบบไม่เป็นเชิงเส้น (Non – Linear)

3.3.1 คุณสมบัติและความสามารถของโครงข่ายประสาทเทียม [8]

คุณสมบัติและความสามารถของโครงข่ายประสาทเทียม สามารถสรุปได้ดังนี้

1. โครงข่ายประสาทเทียมไม่ใช่โปรแกรมคอมพิวเตอร์แต่โครงข่ายจะเรียนรู้จากตัวอย่าง
2. มีความยืดหยุ่นสูงจนสามารถจำลองกระบวนการของปัญหาใดๆได้
3. มีความสามารถในการจำชุดของค่าอินพุต, เอาต์พุตที่มีความซับซ้อนมากจนไม่สามารถจำลองแบบในเชิงความน่าจะเป็นได้
4. มีความสามารถในการปรับตัวเข้ากับการเปลี่ยนแปลงของสิ่งแวดล้อม
5. มีความสามารถในการตอบสนองต่อข้อมูลที่ไม่เคยเห็น
6. ความรู้กระจายอยู่ทั่วโครงสร้างของโครงข่ายประสาทเทียม

3.3.2 ลักษณะของปัญหาที่เหมาะสมกับโครงข่ายประสาทเทียม

ลักษณะของปัญหาที่เหมาะสมกับโครงข่ายประสาทเทียมนี้นี้

1. ตัวอย่างของข้อมูลสำหรับการฝึกสอนประกอบด้วยองค์ประกอบหลายๆค่า
2. เอาต์พุตที่ต้องการสามารถเป็นค่าจำนวนเต็ม หรือจำนวนจริงหรือเป็นเวกเตอร์ของจำนวนเต็มหรือจำนวนจริงก็ได้
3. ตัวอย่างของข้อมูลสามารถมีความผิดพลาดปะปนอยู่ได้
4. ไม่มีปัญหากับขั้นตอนการฝึกหัดซึ่งอาจใช้เวลานานมาก
5. ต้องการประเมินผลเอาต์พุตอย่างรวดเร็ว
6. ไม่ต้องการเข้าใจวิธีการแก้ปัญหา ต้องการให้ปัญหาได้รับการแก้ไขเท่านั้น

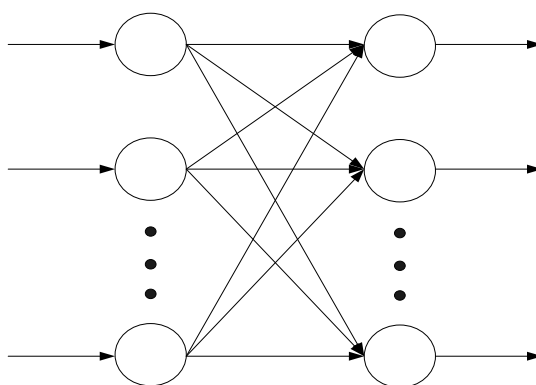
3.4 โครงสร้างของโครงข่ายประสาทเทียม [8]

ในการสร้างโครงข่ายประสาทเทียมมาใช้งานได้นั้น ต้องใช้เซลล์ประสาทเทียมที่มีคุณลักษณะต่างกัน (ค่าถ่วงน้ำหนักจะทำให้คุณสมบัติของเซลล์ประสาทเทียมแต่ละเซลล์มีคุณลักษณะแตกต่างกันไป) มาเชื่อมต่อเข้าด้วยกันให้เกิดลักษณะของโครงข่ายเป็นชั้นๆ ที่เรียกว่า เลเยอร์ (Layer) ซึ่งเซลล์ประสาทแต่ละตัวที่อยู่ในชั้นเดียวกันจะไม่มีการเชื่อมต่อถึงกันประเภทของโครงสร้างการ

เชื่อมต่อโครงข่ายประสาทเทียม พอจะแบ่งออกเป็น 3 แบบ คือ โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Neural Network), โครงข่ายประสาทเทียมแบบหลายชั้น (Multi Layer Neural Network) และโครงข่ายประสาทเทียมแบบแลตทิซ (Lattice Neural Network)

3.4.1 โครงข่ายประสาทเทียมแบบชั้นเดียว (Single Layer Neural Network)

ประกอบด้วยเซลล์ประสาทที่จัดเรียงตัวอยู่ในชั้นต่างๆ ได้แก่ ชั้นอินพุต (Input Layer) และชั้นเอาต์พุต (Output Layer) สาเหตุที่จัดโครงสร้างแบบนี้เป็นแบบชั้นเดียว เนื่องจากเรานับชั้นเอาต์พุตเพียงชั้นเดียวเท่านั้นว่า เป็นชั้นของเซลล์ประสาท สำหรับชั้นอินพุตเราไม่พิจารณาว่าเป็นชั้นของเซลล์ประสาทก็เพราะเป็นชั้นที่ไม่มีการประมวลใดๆ จะทำหน้าที่เพียงแค่รับข้อมูลอินพุตเข้ามาและส่งต่อไปกับชั้นถัดไปเท่านั้น รูปที่ 3.3 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบชั้นเดียว



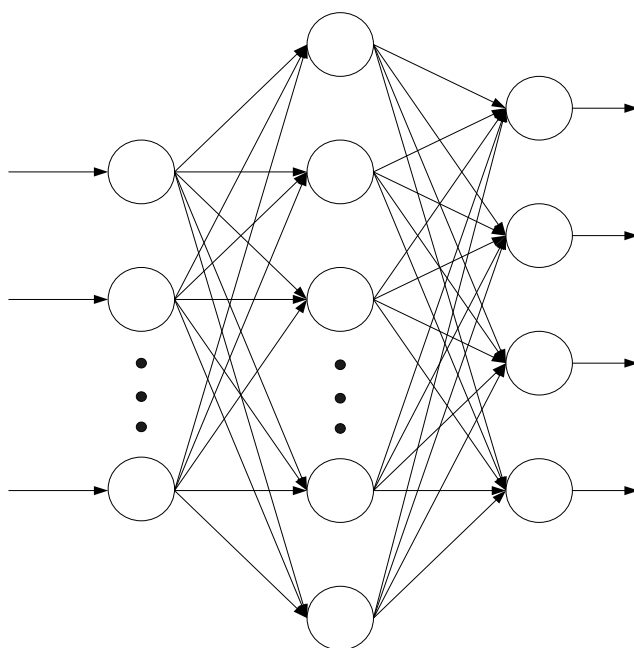
รูปที่ 3.3 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบชั้นเดียว

3.4.2 โครงข่ายประสาทเทียมแบบหลายชั้น (Multi Layer Neural Network)

โครงข่ายประสาทเทียมแบบชั้นเดียวมีข้อจำกัด ไม่สามารถนำมาใช้แก้ปัญหาต่างๆ ในบางกรณีได้ นักวิจัยจึงได้นำเสนอรูปแบบของโครงข่ายประสาทเทียมแบบหลายชั้น ซึ่งมีโครงสร้างประกอบขึ้นจากการจัดเรียงของตัวเซลล์ประสาทตั้งแต่ 2 ชั้นขึ้นไป (ไม่นับรวมชั้นอินพุต) โดยปกติโครงข่ายประสาทเทียมแบบหลายชั้นจะประกอบด้วยโครงข่ายประสาทเทียมแบบชั้นเดียวหลายๆ ตัวต่ออยู่ในโครงสร้างที่มีลำดับชั้นลดหลั่นกันลงไป โดยอยู่ในรูปของการทำงานที่ป้อนไปข้างหน้า (Feed - Forward) ซึ่งชั้นที่อยู่ระหว่างชั้นอินพุตและชั้นเอาต์พุต เรียกว่า ชั้นซ่อน (Hidden Layer) ซึ่งชั้นซ่อนนี้เองที่เป็นตัวเพิ่มความสามารถให้แก่โครงข่าย ทั้งนี้จำนวนของชั้นซ่อนสามารถ

W_k ,

มีได้มากกว่า 1 ชั้น จำนวนของชั้นซ่อนและจำนวนของนิวรอนในแต่ละชั้นจะมีค่าไม่คงที่ ซึ่งแต่ละชั้นอาจจะมีจำนวนนิวรอนที่แตกต่างกันขึ้นอยู่กับการประยุกต์ใช้งาน จากรูปที่ 3.4 แสดงชุดข้อมูลเอาต์พุตที่ได้จากชั้นหนึ่งจะถูกส่งให้เป็นข้อมูลอินพุตสำหรับชั้นต่อไป โดยค่าเอาต์พุตชั้นสุดท้ายซึ่งก็คือ เอาต์พุตของโครงข่ายประสาทเทียมจะถูกนำไปเปรียบเทียบกับค่าความผิดพลาดกับค่าเอาต์พุตที่ต้องการ (Target Output) ค่าความผิดพลาดที่คำนวณได้จะนำไปใช้ในการปรับค่าถ่วงน้ำหนักของโครงข่ายประสาทเทียมให้เปลี่ยนไปในทางที่ถูกต้องมากขึ้น

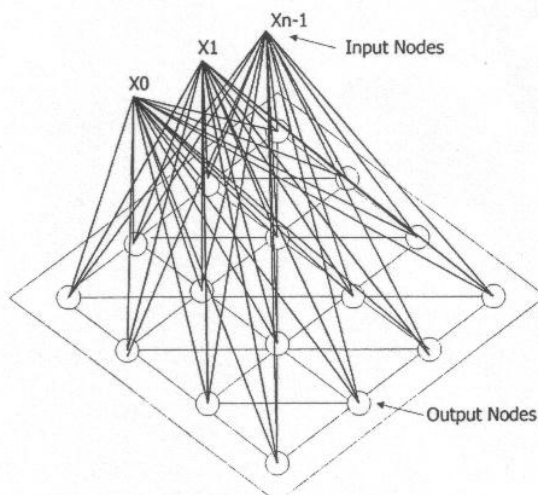

 W_j

รูปที่ 3.4 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบหลายชั้น

3.4.3 โครงข่ายประสาทเทียมแบบแลตทิซ (Lattice Neural Network)

โครงสร้างอาจประกอบด้วยอาร์เรย์ (Array) ของนิวรอนมิติเดียวหรือหลายมิติก็ได้ มิติโครงสร้างแบบแลตทิซคือมิติของการจัดเรียงนิวรอน การจัดโครงสร้างแบบมิติเดียวคือการจัดเรียงนิวรอนในแนวนอนหรือแนวตั้งอย่างใดอย่างหนึ่งเพียงแถวเดียว การจัดโครงสร้างแบบ 2 มิติเป็นการจัดเรียงนิวรอนทั้งในแนวนอนและแนวตั้ง ดังตัวอย่างโครงสร้างแบบ 2 มิติที่มีขนาดนิวรอน 4×4 ในรูปที่ 3.5

 X_p



รูปที่ 3.5 แสดงโครงสร้างของโครงข่ายประสาทเทียมแบบแลตทิซ

3.5 การเรียนรู้ของโครงข่ายประสาทเทียม

เมื่อสร้างโครงข่ายประสาทเทียมจะต้องมีการฝึกสอนเพื่อให้มีคุณลักษณะตามต้องการ ซึ่งการเรียนรู้ของโครงข่ายประสาทเทียมจะมีประสิทธิภาพเพียงใดนั้นขึ้นอยู่กับค่าถ่วงน้ำหนักของโครงข่ายประสาทเทียม ซึ่งการสอนโครงข่ายประสาทเทียมก็คือการหาค่าถ่วงน้ำหนักที่เหมาะสมให้แก่โครงข่ายนั้นๆ ซึ่งโครงข่ายประสาทเทียมสามารถแบ่งออกเป็น 2 ประเภทตามลักษณะการเรียนรู้คือ การเรียนรู้แบบควบคุม (Supervised Learning) และการเรียนรู้ไม่มีการควบคุม (Unsupervised Learning)

3.5.1 การเรียนรู้แบบควบคุม (Supervised Learning)

การเรียนรู้วิธีนี้จะมีข้อมูลที่เราเรียกว่า ชุดฝึกสอน (Training Pattern) โดยในชุดฝึกสอนนี้จะมีข้อมูลอินพุตและเอาต์พุต ซึ่งโครงข่ายประสาทเทียมจะถูกฝึกสอนจากข้อมูลในชุดฝึกสอนเพื่อหาค่าถ่วงน้ำหนักที่เหมาะสม โดยนำเอาต์พุตที่ได้จากโครงข่ายประสาทเทียมไปเปรียบเทียบกับเอาต์พุตของชุดฝึกสอน จากนั้นค่าผิดพลาดที่ได้จากความแตกต่างของเอาต์พุตจากโครงข่ายประสาทเทียมและเอาต์พุตจากชุดฝึกสอนก็จะถูกป้อนกลับไปใหม่ เพื่อให้คำนวณหาค่าถ่วงน้ำหนักใหม่ เพื่อให้ค่าความผิดพลาดระหว่างเอาต์พุตของโครงข่ายประสาทเทียมกับเอาต์พุตของชุดฝึกสอนมีค่าน้อยที่สุด

3.5.2 การเรียนรู้แบบไม่มีการควบคุม (Unsupervised Learning)

การเรียนรู้วิธีนี้จะเป็นการเรียนรู้ที่ไม่ต้องอาศัยชุดฝึกสอนมาทำการฝึกสอนให้กับโครงข่ายประสาทเทียม การปรับค่าถ่วงน้ำหนักนั้นนิรอนจะทำการวิเคราะห์ด้วยตัวของมันเอง นั่นคือการ

ปรับค่าถ่วงน้ำหนักจะไม่มีการกำหนดค่าที่ถูกต้องว่าควรจะใช้ค่าถ่วงน้ำหนักเท่าไร แต่นิวรอนจะอาศัยการป้อนกลับไปของค่าผิดพลาดและวิเคราะห์หาค่าถ่วงน้ำหนักด้วยตัวของมันเอง

3.6 ฟังก์ชันกระตุ้น (Activation function) [10]

ฟังก์ชันกระตุ้นหรือหนังสือบางเล่มเรียกว่า ฟังก์ชันการถ่ายโอน (Transfer Function) เป็นตัวกำหนดค่าเอาต์พุต ฟังก์ชันกระตุ้นที่ใช้ในการฝึกสอนโครงข่ายประสาทเทียมมีอยู่หลายชนิด เราสามารถแบ่งฟังก์ชันกระตุ้นพื้นฐานออกเป็น 3 ชนิดคือ

1. ฟังก์ชันเทรชโฮลด์ (Threshold Function หรือ Hard Limit Function หรือ Step Function)
2. ฟังก์ชันเชิงเส้นอิ่มตัว (Saturating Linear Function หรือ Piecewise – Linear Function)
3. ฟังก์ชันซิกมอยด์ (Sigmoid Function)

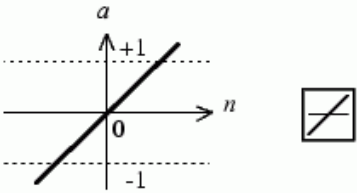
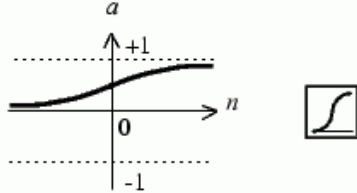
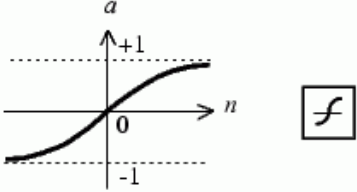
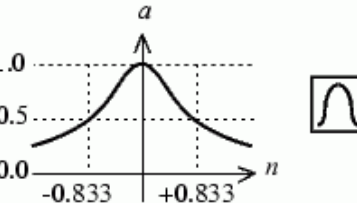
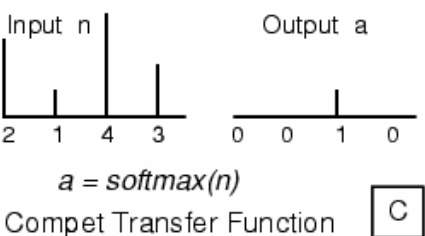
จากฟังก์ชันพื้นฐานทั้ง 3 ชนิด ฟังก์ชันซิกมอยด์เป็นฟังก์ชันที่ได้รับความนิยมนำมาใช้เป็นฟังก์ชันกระตุ้นของโครงข่ายประสาทเทียมกันมาก เนื่องจากเป็นฟังก์ชันที่เพิ่มขึ้นอย่างสม่ำเสมอ (Smoothness) นอกเหนือจากฟังก์ชันกระตุ้นพื้นฐานทั้งสาม แล้วยังสามารถนำความสัมพันธ์ระหว่างชุดข้อมูลในรูปอื่นๆมาเป็นฟังก์ชันกระตุ้นได้ ตัวอย่างเช่น ฟังก์ชันกระตุ้นแบบคอมเพทิทีฟ (Competitive Function)

โดยทั่วไปฟังก์ชันกระตุ้นจะจำกัดช่วงเอาต์พุตอยู่ตั้งแต่ 0 ถึง 1 ในการประยุกต์ใช้งานบางกรณีอาจมีความจำเป็นต้องใช้ฟังก์ชันที่มีค่าออกมานอกเหนือช่วงดังกล่าว ตัวอย่างเช่น ต้องการเอาต์พุตอยู่ในช่วงตั้งแต่ -1 ถึง +1 สำหรับการพิจารณาเลือกใช้ฟังก์ชันกระตุ้นแบบใดนั้นนับถึงปัจจุบันยังไม่มีข้อกำหนดที่แน่นอน แต่อาจจะพิจารณาองค์ประกอบดังต่อไปนี้

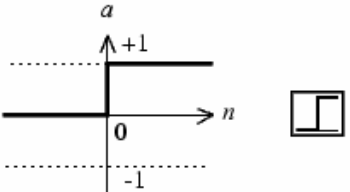
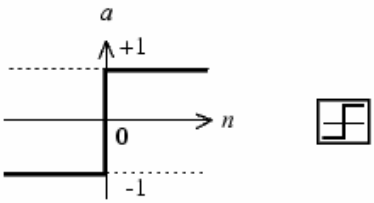
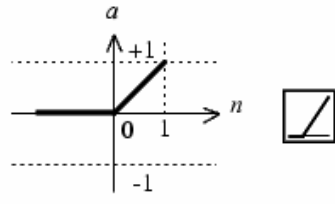
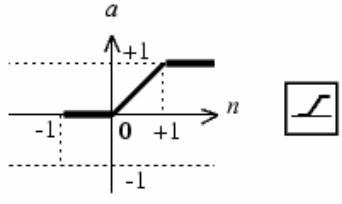
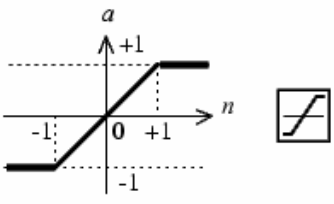
1. ความเป็นเชิงเส้นหรือไม่เป็นเชิงเส้น
2. ลักษณะของข้อมูลอินพุต
3. ขอบเขตที่ต้องการและลักษณะของข้อมูลเอาต์พุต
4. ปัจจัยอื่นที่มีนัยสำคัญต่อผลตอบแทนของนิวรอนหรือโครงข่ายประสาทเทียมโดยรวม

ในบางครั้งอาจเลือกใช้ฟังก์ชันสโตกาสติก (Stochastic Function) เป็นฟังก์ชันกระตุ้นในกรณีค่าความน่าจะเป็นของข้อมูลอินพุตมีนัยสำคัญ หรือในบางกรณีอาจเลือกใช้ฟังก์ชันแบบเชิงเส้นและไม่เป็นเชิงเส้นในโครงข่ายประสาทเทียมเดียวกัน ตัวอย่างเช่น โครงข่ายประสาทเทียมชนิดฟังก์ชันฐานราศมี (RBF) ซึ่งใช้ฟังก์ชันกระตุ้นแบบไม่เป็นเชิงเส้นในชั้นซ่อนของโครงข่ายประสาท ขณะที่ชั้นเอาต์พุตใช้ฟังก์ชันกระตุ้นแบบเป็นเชิงเส้น ตารางที่ 3.1 แสดงฟังก์ชันกระตุ้นที่ใช้ในการฝึกสอนโครงข่ายประสาทเทียม

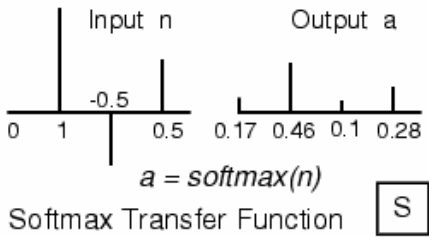
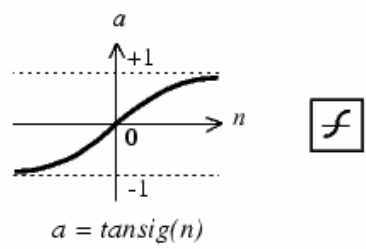
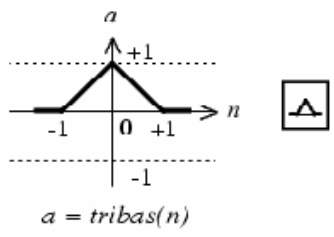
ตารางที่ 3.1 แสดงชนิดฟังก์ชันการถ่ายโอน

ชนิดฟังก์ชันการถ่ายโอน	สมการฟังก์ชันการถ่ายโอน
 $a = \text{purelin}(n)$ Linear Transfer Function	$a(n) = n$
 $a = \text{logsig}(n)$ Log-Sigmoid Transfer Function	$a(n) = \frac{1}{1 + e^{-n}}$
 $a = \text{tansig}(n)$ Tan-Sigmoid Transfer Function	$a(n) = \frac{1 - e^{(-n)}}{1 + e^{(-n)}}$
 $a = \text{radbas}(n)$ Radial Basis Function	$a(n) = e^{-n^2}$
 $a = \text{softmax}(n)$ Compet Transfer Function	$a(n_i) = \begin{cases} 1; & i = i^* \\ 0; & i \neq i^* \end{cases}$

ตารางที่ 3.1 (ต่อ)

ชนิดฟังก์ชันการถ่ายโอน	สมการฟังก์ชันการถ่ายโอน
 $a = \text{hardlim}(n)$ Hard-Limit Transfer Function	$a(n) = \begin{cases} 1; & n \geq 0 \\ 0; & n < 0 \end{cases}$
 $a = \text{hardlims}(n)$ Symmetric Hard-Limit Trans. Funct.	$a(n) = \begin{cases} 1; & n \geq 0 \\ -1; & n < 0 \end{cases}$
 $a = \text{poslin}(n)$ Positive Linear Transfer Funct.	$a(n) = \begin{cases} n; & n \geq 0 \\ 0; & n < 0 \end{cases}$
 $a = \text{satlin}(n)$ Satlin Transfer Function	$a(n) = \begin{cases} 0; & n \leq 0 \\ n; & 0 \leq n \leq 1 \\ 1; & 1 \leq n \end{cases}$
 $a = \text{satlins}(n)$ Satlins Transfer Function	$a(n) = \begin{cases} -1; & n \leq -1 \\ n; & -1 \leq n \leq 1 \\ 1; & 1 \leq n \end{cases}$

ตารางที่ 3.1 (ต่อ)

ชนิดฟังก์ชันการถ่ายโอน	สมการฟังก์ชันการถ่ายโอน
 $a = \text{softmax}(n)$ Softmax Transfer Function S	$a(n) = \frac{e^n}{\sum e^n}$
 $a = \text{tansig}(n)$ Tan-Sigmoid Transfer Function f	$a(n) = \frac{1 - e^{(-2n)}}{1 + e^{(-2n)}}$
 $a = \text{tribas}(n)$ Triangular Basis Function △	$a(n) = \begin{cases} 1 - n ; & -1 \leq n \leq 1 \\ 0; & \text{otherwise} \end{cases}$

3.7 การเลือกโครงข่ายประสาทเทียม [9]

เพราะทุกโครงข่ายประสาทเทียมมีพื้นฐานอยู่บนหลักการของนิวรอน การเชื่อมต่อและฟังก์ชันถ่ายโอน ทำให้มันมีความแตกต่างระหว่างโครงข่ายหรือสถาปัตยกรรมของโครงข่ายประสาท ความหลากหลายเกิดจากกฎการเรียนรู้ต่างๆ และวิธีการที่กฎต่างๆ เหล่านั้นปรับเปลี่ยนโทโปโลยีของโครงข่าย

โดยพื้นฐาน การใช้งานโครงข่ายประสาทส่วนมากจะอยู่ในประเภททั้ง 5 นี้

1. Prediction
2. Classification
3. Data association
4. Data conceptualization
5. Data filtering

ตารางที่ 3.2 แสดงการเลือกโครงข่ายแต่ละประเภท

ประเภทของโครงข่าย	โครงข่าย	การใช้งาน
Prediction	- Back-propagation - Delta Bar Delta - Extended delta bar delta - Directed random search - Higher order Neural Networks - Self Organizing Map into Backpropagation	ใช้ค่าของอินพุตเพื่อทำนายหาเอาต์พุต (เช่น เลือกหุ้นที่ดีที่สุดในตลาดหุ้น, การพยากรณ์อากาศ, การระบุบุคคลที่เสี่ยงต่อโรคมะเร็ง)
Classification	- Learning vector quantization - Counter-propagation - Probabalistic neural network	ใช้ค่าของอินพุตเพื่อกำหนดการจัดหมวดหมู่ (เช่น อินพุตใช้อักษร A หรือไม่)
Data association	- Hopfield - Boltzmann Machine - Hamming network - Bidirectional associative memory - Spatio-temporal pattern recognition	เหมือนกับ Classification แต่จะสามารถจดจำข้อมูลที่มีความผิดพลาดได้ (เช่น ไม่เพียงแต่ระบุอักษรที่ได้แสดกน สามารถระบุได้เมื่อเครื่องแสดกนทำงานไม่ถูกต้อง)
Data conceptualization	- Adaptive resonance Network - Self organizing map	วิเคราะห์อินพุตเพื่อให้สามารถสรุปความสัมพันธ์ของการจัดกลุ่มได้ (เช่น ดึงรายชื่อจากฐานข้อมูล ที่น่าจะซื้อสินค้าตัวหนึ่งมากที่สุด)
Data filtering	- Recirculation	ทำให้สัญญาณอินพุตเรียบขึ้น (เช่น กำจัดสัญญาณรบกวน)

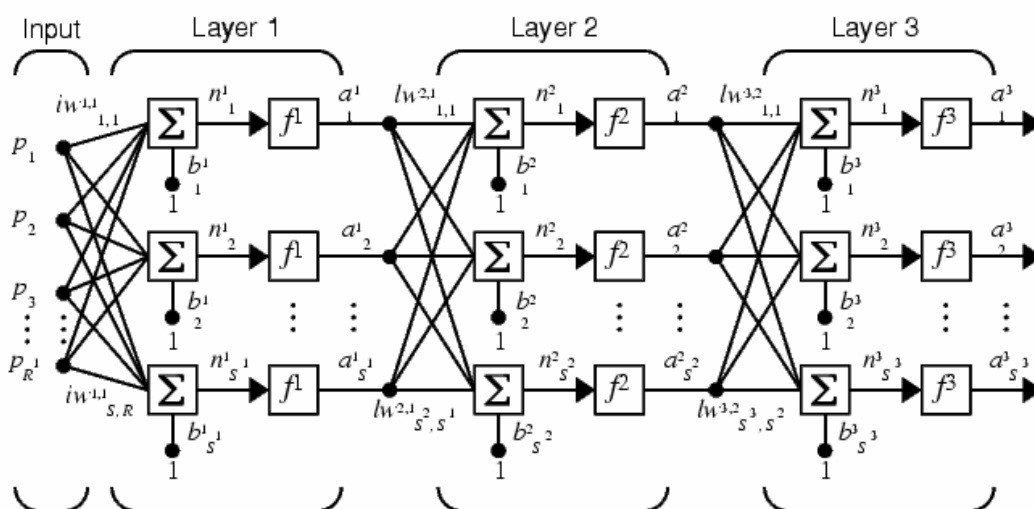
ตารางที่ 3.2 แสดงความแตกต่างระหว่างโครงข่ายประเภทต่างๆ และแสดงโทโปโลยีต่างๆ ของโครงข่ายประเภทนั้นๆ ตารางนี้เป็นเพียงการแนะแนวทางส่วนหนึ่งเท่านั้น ไม่ใช่ทั้งหมด บางโครงข่ายที่ถูกจัดอยู่ในกลุ่มอาจใช้แก้ปัญหาได้มากกว่าหนึ่งประเภท Feed forward back-

propagation ถูกใช้เพื่อแก้ปัญหาเกือบทุกประเภท และเป็นที่ยอมรับมากที่สุดสำหรับปัญหา 4 ประเภทแรก

3.8 โครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ (Back – Propagation Neural Network) [10]

3.8.1 โครงสร้าง

โครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับ (Back – Propagation Neural Network : BP) ประกอบด้วยนิวรอนที่เรียงกันอยู่มากกว่า 2 ชั้น คือมีได้ตั้งแต่ 3 ชั้นขึ้นไป ซึ่งประกอบด้วยชั้นอินพุต, ชั้นซ่อนและชั้นเอาต์พุต โครงข่ายชนิดนี้มีลักษณะเป็นโครงข่ายที่ป้อนไปข้างหน้า มีการเชื่อมโยงในแต่ละชั้นแบบต่อถึงกันหมด นั่นคือ ทุกๆนิวรอนในชั้นอินพุตจะส่งสัญญาณไปยังทุกนิวรอนในชั้นซ่อนชั้นแรก และทุกนิวรอนในชั้นซ่อนชั้นแรกจะส่งสัญญาณไปยังทุกนิวรอนในชั้นถัดไปจนในที่สุดทุกนิวรอนในชั้นสุดท้ายจะส่งสัญญาณไปยังทุกนิวรอนในชั้นเอาต์พุต ดังรูปที่ 3.6 แสดงโครงสร้างพื้นฐานของโครงข่าย 3 ชั้น



รูปที่ 3.6 แสดงแบบจำลองของโครงข่ายประสาทเทียมที่มี 2 ชั้นซ่อน[10]

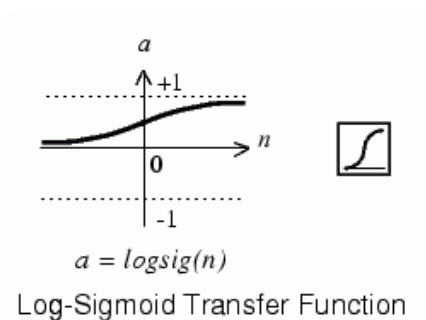
จากรูปที่ 3.6 โครงข่ายประสาทเทียมมี 2 ชั้นซ่อน มีอินพุต R ชุด มีนิวรอนในชั้นที่ 1 และ 2 เท่ากับ S^1 และ S^2 และมีเอาต์พุต S^3 ชุด แต่ละสมาชิกของอินพุตเวกเตอร์ P จะถูกเชื่อมเข้ากับนิวรอนแต่ละนิวรอนในชั้นซ่อนที่ 1 ด้วยค่าถ่วงน้ำหนัก W^1 (เรียกว่า Input Weight, $IW^{1,1}$) ซึ่งเป็นเมตริกซ์ขนาด $S^1 \times R$ และเวกเตอร์ a^1 ซึ่งผลลัพธ์ของชั้นซ่อนที่ 1 และขนาด S^1 จะเป็นอินพุตของชั้นซ่อนที่ 2 สมาชิกแต่ละตัวของเวกเตอร์ a^1 ก็จะถูกเชื่อมเข้ากับนิวรอนแต่ละนิวรอนในชั้นซ่อนที่ 2

ด้วยค่าถ่วงน้ำหนัก W^2 (เรียกว่า Layer Weight, $LW^{2,1}$) ซึ่งเป็นเมตริกซ์ขนาด $S^2 \times S^1$ เช่นเดียวกับชั้นซ่อนที่ 1 เวกเตอร์ a^2 ซึ่งผลลัพธ์ของชั้นซ่อนที่ 2 และขนาด S^2 จะเป็นอินพุตของชั้นข้อมูลเอาต์พุต สมาชิกแต่ละตัวของเวกเตอร์ a^2 จะถูกเชื่อมเข้ากับนิวรอนแต่ละตัวชั้นข้อมูลเอาต์พุตด้วยค่าถ่วงน้ำหนัก W^3 (เรียกว่า $LW^{3,2}$) ซึ่งเป็นเมตริกซ์ขนาด $S^3 \times S^2$ โดยที่เวกเตอร์ a^3 ซึ่งเป็นเมตริกซ์ขนาด S^3 จะเป็นผลลัพธ์ของโครงข่ายประสาทเทียม

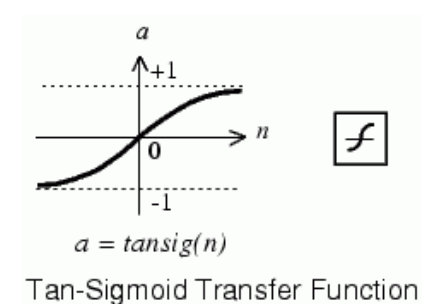
ฟังก์ชันการถ่ายโอนที่ใช้ในโครงข่ายประสาทเทียมแบบแพร่ค่าย้อนกลับนี้ได้แก่ Log – Sigmoid Transfer Function และ Tan – Sigmoid Transfer Function ดังรูปที่ 3.7 และ 3.8 ซึ่งมีสมการเป็นดังสมการที่ 3.1 และ 3.2 ตามลำดับ

$$a(n) = \frac{1}{1 + e^{-n}} \quad (3.1)$$

$$a(n) = \frac{1 - e^{(-n)}}{1 + e^{(-n)}} \quad (3.2)$$



รูปที่ 3.7 แสดงฟังก์ชันถ่ายโอน Log – Sigmoid



รูปที่ 3.8 แสดงฟังก์ชันถ่ายโอน Tan – Sigmoid

3.8.2 การทำงาน

ขั้นตอนการฝึกสอนของโครงข่ายประสาทเทียมชนิดแพร่ค่าย้อนกลับสามารถแบ่งออกเป็น 3 ขั้นตอนด้วยกัน คือ ป้อนข้อมูลอินพุตไปข้างหน้า, แพร่ค่าย้อนกลับของค่าผิดพลาดที่เกี่ยวข้องและปรับค่าถ่วงน้ำหนัก

ระหว่างป้อนไปข้างหน้า จะเริ่มต้นด้วยการนำค่าอินพุตที่ได้รับจากชั้นอินพุตมาคำนวณหาผลรวมที่ได้รับในแต่ละนิวรอน แล้วจึงส่งผลลัพธ์ที่ได้ส่งต่อไปยังชั้นซ่อนที่ 1 แต่ละนิวรอนในชั้นซ่อนที่ 1 จะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นนั้น แล้วจึงส่งผลลัพธ์ที่ได้ส่งไปยังชั้นซ่อนที่อยู่ถัดไป ในชั้นซ่อนที่ 2 แต่ละนิวรอนในชั้นซ่อนที่ 2 จะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นนั้น จากนั้นจะส่งผลลัพธ์ที่ได้ส่งต่อไปยังชั้นถัดไป สุดท้ายแต่ละนิวรอนในชั้นเอาต์พุตจะทำการคำนวณด้วยฟังก์ชันกระตุ้นในชั้นชั้นเอาต์พุต เพื่อหาคำตอบหรือผลตอบสนองของโครงข่ายสำหรับแต่ละรูปแบบข้อมูลที่ป้อนให้ ทิศทางการไหลของข้อมูลเหล่านี้จะเป็นลักษณะเคลื่อนไปข้างหน้าจากชั้นของอินพุต ผ่านชั้นของนิวรอนเซลล์ภายในไปสู่ชั้นเอาต์พุต วิธีการคำนวณของโครงข่ายลักษณะนี้ จึงมีชื่อตามลักษณะดังกล่าวว่า Forward – Propagation

ขั้นต่อไปจะเป็นการคำนวณแบบ Backward – Propagation สำหรับขั้นตอนนี้จะเป็นการนำผลลัพธ์ที่ได้จากเอาต์พุตของโครงข่ายมาเปรียบเทียบกับเอาต์พุตเป้าหมายเพื่อคำนวณค่าผิดพลาดที่เกี่ยวข้อง แล้วค่าผลต่างระหว่างเอาต์พุตของโครงข่ายกับเอาต์พุตเป้าหมายที่ได้จากโครงข่ายจะถูกส่งถอยหลังกลับมายังชั้นซ่อนที่ 2 และส่งต่อไปยังนิวรอนต่างๆภายในชั้นซ่อนที่ 2 แต่ละโหนดจะได้รับค่าผิดพลาดเพียงบางส่วนขึ้นอยู่กับว่านิวรอนเซลล์นั้นเป็นตัวส่งผลมากหรือน้อยไปสู่เอาต์พุตนั้น เพื่อปรับค่าถ่วงน้ำหนักระหว่างชั้นเอาต์พุตกับชั้นซ่อนที่ 2 กระบวนการของการส่งค่าความผิดพลาดกลับมานั้นจะซ้ำกับชั้นถัดลงมาอีกจนกระทั่งทุกโหนดในโครงข่ายได้รับส่วนแบ่งค่าความผิดพลาดนั้น

ขั้นตอนสุดท้ายจะนำค่าผิดพลาดที่ได้มาใช้ในการปรับค่าถ่วงน้ำหนัก ซึ่งค่าถ่วงน้ำหนักจะเปลี่ยนไปมากหรือน้อยขึ้นอยู่กับปริมาณของค่าความผิดพลาดที่ได้รับ กระบวนการจะกลับไปทำซ้ำในขั้นตอนแรกจนกระทั่งค่าความผิดพลาดที่ได้มีค่าน้อยกว่าค่าผิดพลาดค่าสุดท้ายที่กำหนดไว้

โครงข่ายประสาทเทียมแบบแพร่ค่าย้อนกลับนั้นจะมีจำนวนนิวรอนในชั้นอินพุตและชั้นเอาต์พุตคงที่ ซึ่งเท่ากับจำนวนอินพุตและเอาต์พุตของปัญหานั้นเอง ส่วนจำนวนนิวรอนในชั้นซ่อนนั้นสามารถปรับเปลี่ยนได้ตามความเหมาะสมและไม่มีจำนวนที่เป็นสูตรสำเร็จ ดังนั้นในการฝึกสอนของโครงข่ายประสาทเทียมจึงต้องมีการปรับเปลี่ยนจำนวนนิวรอนในชั้นซ่อนเพื่อหาโครงสร้างที่เหมาะสมเพื่อให้ค่าผิดพลาดระหว่างคำตอบที่แท้จริงกับผลลัพธ์ที่ได้จากโครงข่ายประสาทเทียมนั้นมีค่าน้อยที่สุด จำนวนนิวรอนในชั้นซ่อนนั้นมีผลต่อระยะเวลาในการฝึกสอนด้วย นั่นคือถ้ามีจำนวนนิวรอนในชั้นซ่อนมากเวลาที่ใช้ในการฝึกสอนก็จะมากกว่าโครงข่ายประสาทเทียมที่มีจำนวนนิวรอนในชั้นซ่อนน้อยกว่า เมื่อฝึกสอนด้วยเงื่อนไขเดียวกัน แต่อย่างไรก็ตามจำนวน

นิวรอนในชั้นซ่อนที่มากก็ไม่ได้เป็นตัวบ่งบอกว่าโครงข่ายจะมีประสิทธิภาพมากกว่าโครงข่ายประสาทเทียมที่จำนวนนิวรอนในชั้นซ่อนที่น้อยกว่า ซึ่งต้องทำการลองปรับเปลี่ยนดูเพื่อหาโครงข่ายที่เหมาะสมที่สุด

3.9 โครงข่ายประสาทเทียมชนิด Generalized Regression [10]

โครงข่ายประสาทเทียมแบบ Generalized Regression นั้นเป็นโครงข่ายประสาทเทียมที่มีลักษณะคล้ายกับโครงข่ายประสาทเทียมแบบฟังก์ชันฐานรัศมี (Radial Basis Function Neural Network : RBF)

3.9.1 Radial Basis Function [11]

แนวคิดของเน็ตเวิร์ค Radial Basis Function (RBF) ได้มาจากทฤษฎีของการประมาณฟังก์ชัน (function approximation) เน็ตเวิร์ค RBF ใช้วิธีการที่แตกต่างจากเน็ตเวิร์ค Multi-Layer Perceptron (MLP) ในการเรียนรู้เพื่อประมาณฟังก์ชัน คุณลักษณะหลัก ได้แก่

1. เป็นเน็ตเวิร์คแบบ feed-forward ที่มี 2 เลเยอร์
2. โหนดในชั้นซ่อน (hidden node) สร้างชุดของฟังก์ชัน radial basis (เช่น Gaussian function)
3. โหนดในชั้นข้อมูลด้านออก (output node) สร้างฟังก์ชันผลรวมเชิงเส้นเหมือนใน MLP
4. การฝึกสอนเน็ตเวิร์คถูกแบ่งออกเป็นสองระยะ : ระยะแรกจะหาค่าถ่วงน้ำหนัก (weight) จากชั้นข้อมูลด้านเข้าไปยังชั้นซ่อน ระยะที่สองจะหาค่าถ่วงน้ำหนัก (weight) จากชั้นซ่อนไปยังชั้นข้อมูลด้านออก
5. การฝึกสอนและการเรียนรู้รวดเร็วมาก
6. เป็นเน็ตเวิร์คที่ดี/เหมาะสมสำหรับการสอดแทรก (interpolation)

3.9.1.1 ฟังก์ชัน basis ที่ใช้โดยทั่วไป

1. Gaussian Function:

$$a(n) = e^{-n^2} \quad (3.3)$$

2. Thin Plate Spline Function:

$$a(n) = n^2 \log(n) \quad (3.4)$$

3. Multi-Quadric Function:

$$a(n) = (n^2 + \alpha^2)^{1/2} \quad \alpha > 0 \quad (3.5)$$

4. Inverse Multi-Quadric Function:

$$a(n) = (n^2 + \alpha^2)^{-1/2} \quad \alpha > 0 \quad (3.6)$$

5. Generalized Multi-Quadric Function:

$$a(n) = (n^2 + \alpha^2)^\beta \quad \alpha > 0, \quad 1 > \beta > 0 \quad (3.7)$$

6. Generalized Inverse Multi-Quadric Function:

$$a(n) = (n^2 + \alpha^2)^{-\delta} \quad \alpha > 0, \quad \delta > 0 \quad (3.8)$$

7. Cubic Function:

$$a(n) = n^3 \quad (3.9)$$

8. Linear Function:

$$a(n) = n \quad (3.10)$$

3.9.1.2 คุณสมบัติของฟังก์ชัน Radial Basis

ฟังก์ชัน Gaussian และ Inverse Multi-Quadric ถูกจำกัดวงในความหมายที่

$$a(n) \rightarrow 0 \quad \text{เมื่อ} \quad |n| \rightarrow \infty$$

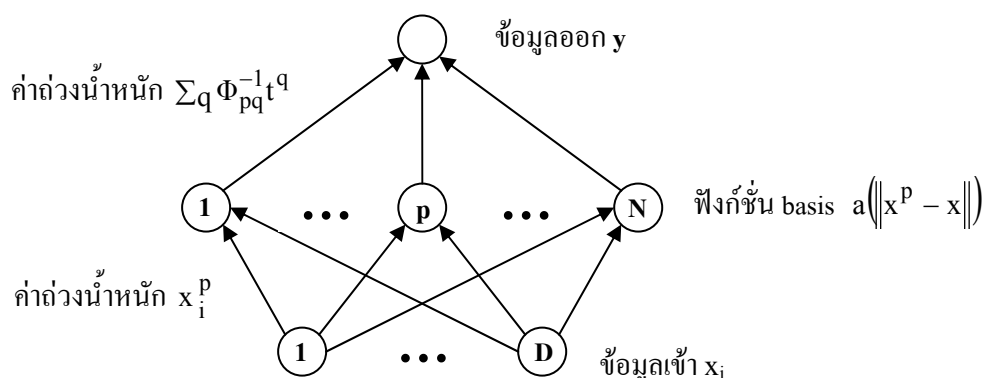
ส่วนฟังก์ชันอื่นๆ มีคุณสมบัติที่

$$a(n) \rightarrow \infty \quad \text{เมื่อ} \quad |n| \rightarrow \infty$$

สำหรับการทำโครงข่ายประสาท มีเหตุผลที่ดีสำหรับการเลือกฟังก์ชัน basis ที่ถูกจำกัดวง อย่างเช่นฟังก์ชัน Gaussian basis เพราะมีคุณสมบัติอื่นๆ ในด้านการวิเคราะห์ที่เป็นประโยชน์

3.9.1.3 เน็ตเวิร์คของฟังก์ชัน Radial Basis

โครงข่ายประสาทจะมีลักษณะเป็นดังรูปที่ 3.9



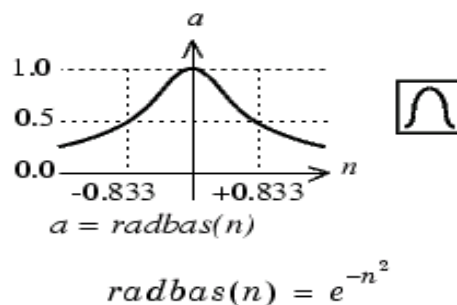
รูปที่ 3.9 แสดงโครงข่ายที่ใช้ฟังก์ชัน Radial Basis [11]

หมายเหตุ รูปแบบของการฝึกสอน $\{x_i^p, t^p\}$ จำนวน N รูปแบบเป็นตัวกำหนดค่าถ่วงน้ำหนักโดยตรง ค่าถ่วงน้ำหนักจากชั้นซ่อนไปชั้นข้อมูลด้านออกจะคูณการกระตุ้นของหน่วยชั้นซ่อน (hidden unit) ในวิธีธรรมดา แต่ค่าถ่วงน้ำหนักจากชั้นข้อมูลด้านเข้าไปชั้นซ่อนจะถูกใช้ในรูปแบบที่ต่างกันอย่าง

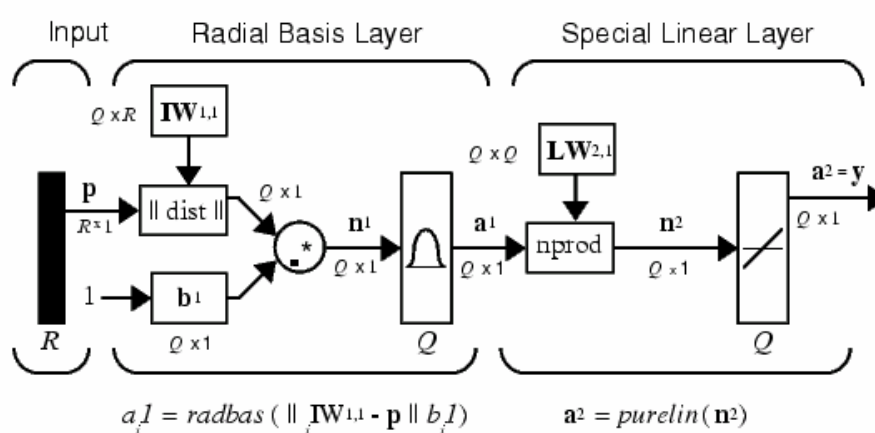
3.9.2 โครงสร้าง [10]

โครงสร้างพื้นฐานของโครงข่ายประสาทเทียมแบบ Generalized Regression (Generalized Regression Neural Network : GRNN) นั้นจะประกอบไปด้วยชั้นของนิวรอนทั้งหมด 3 ชั้น คือ ชั้นอินพุต, ชั้นซ่อนหรือบางครั้งเรียกว่าชั้นฐานรัศมี (Radial Basis Layer) และชั้นเอาต์พุตหรือบางครั้งเรียกว่าชั้นเชิงเส้นพิเศษ (Special Linear Layer) โครงสร้างประสาทเทียมชนิดนี้เป็นโครงข่ายที่ป้อนไปข้างหน้า โดยมีการเชื่อมโยงในแต่ละชั้นแบบต่อถึงกันหมดนั่นคือ ทุกนิวรอนในชั้นอินพุตจะส่งสัญญาณไปทุกๆนิวรอนในชั้นฐานรัศมี และทุกๆนิวรอนในชั้นฐานรัศมีจะส่งสัญญาณไปทุกๆนิวรอนในชั้นเอาต์พุต และใช้ฟังก์ชันถ่ายโอนที่ใช้ในชั้นฐานรัศมีเป็นฟังก์ชันถ่ายโอนชนิด Radial basis เท่านั้น ดังแสดงในรูปที่ 3.10 โดยมีสมการของฟังก์ชันกระตุ้นเป็นดังสมการที่ 3.11 ส่วนในชั้นเอาต์พุตเป็นฟังก์ชันถ่ายโอนชนิด Linear

$$a(n) = e^{-n^2} \quad (3.11)$$



รูปที่ 3.10 แสดงฟังก์ชันถ่ายโอนชนิด Radial Basis Function



รูปที่ 3.11 แสดงแบบจำลองโครงข่ายประสาทเทียมชนิด Generalized Regression [10]

รูปที่ 3.11 แสดงโครงสร้างของโครงข่ายประสาทเทียมชนิดนี้ มีอินพุต R ชุด มีจำนวนนิวรอนในชั้นฐานรัศมีเท่ากับ Q และมีเอาต์พุต Q ชุด แต่ละสมาชิกของอินพุตเวกเตอร์ $\mathbf{p}$ จะถูกเชื่อมเข้ากับนิวรอนแต่ละนิวรอนในชั้นฐานรัศมีด้วยค่าถ่วงน้ำหนัก $\mathbf{W}^1$ ($\text{IW}^{1,1}$) ซึ่งเป็นเมตริกซ์ขนาด $Q \times R$ และเวกเตอร์ $\mathbf{a}^1$ เป็นผลลัพธ์ของชั้นฐานรัศมี โดยมีขนาด Q จะเป็นอินพุตของชั้น Special Linear Layer ด้วยค่าถ่วงน้ำหนัก $\mathbf{W}^2$ ($\text{IW}^{2,1}$) ซึ่งเป็นเมตริกซ์ขนาด $Q \times Q$ โดยเวกเตอร์ $\mathbf{a}^2$ ซึ่งเป็นเมตริกซ์ขนาด Q เป็นผลลัพธ์ของโครงข่ายประสาทเทียม

3.9.3 การทำงาน

การฝึกสอนของโครงข่ายประสาทเทียมชนิด Generalized Regression สามารถแบ่งออกเป็น 2 ขั้นตอนด้วยกัน คือ ป้อนข้อมูลอินพุตไปข้างหน้าและกระบวนการปรับค่าถ่วงน้ำหนัก ซึ่งระหว่างการป้อนไปข้างหน้าจะเริ่มต้นด้วยการนำค่าอินพุตที่ได้รับจากชั้นอินพุตมาคำนวณหาระยะห่างกับค่าถ่วงน้ำหนักในแต่ละนิวรอน แล้วจึงส่งต่อผลลัพธ์ที่ได้ส่งต่อไปคำนวณในฟังก์ชัน

กระตุ้นชนิดฐานรัศมี (Radial Basis) จากนั้นจะทำการส่งผลลัพธ์ที่ได้ไปยังชั้นเอาต์พุตซึ่งอยู่ถัดไป ซึ่งในชั้นเอาต์พุตแต่ละนิวรอน จะทำการจัดระเบียบข้อมูล (Normalization) ผลลัพธ์ที่ได้รับในแต่ละนิวรอน แล้วจึงส่งผลลัพธ์ที่ได้ส่งไปคำนวณกับฟังก์ชันกระตุ้นชนิดเชิงเส้น เพื่อหาคำตอบหรือผลตอบสนองของโครงข่ายประสาทเทียมสำหรับแต่ละรูปแบบข้อมูลที่ป้อนให้ ทิศทางการไหลของข้อมูลเหล่านี้เป็นลักษณะเคลื่อนที่ไปข้างหน้า

การปรับค่าถ่วงน้ำหนักจะปรับโดยเพิ่มจำนวนรอบที่ล้นรอบ และดำเนินการปรับค่าถ่วงน้ำหนักเช่นนี้ ไปเรื่อยๆ จนกระทั่งค่าความผิดพลาดต่ำกว่าที่กำหนดไว้ค่าหนึ่งแล้วจึงหยุดกระบวนการเรียนรู้ ในการฝึกสอนโครงข่ายประสาทเทียมชนิด Generalized Regression นอกจากนี้ จะปรับค่าถ่วงน้ำหนักโดยเพิ่มจำนวนรอบแล้ว ยังสามารถทำการเพิ่มค่าการกระจายของฟังก์ชันกระตุ้นชนิดรัศมี (spread) ควบคู่กันไป เพื่อค่าถ่วงน้ำหนักและค่าไบแอสที่ได้จะเป็นค่าที่เหมาะสมสำหรับทุกคู่ของอินพุตและค่าเป้าหมาย