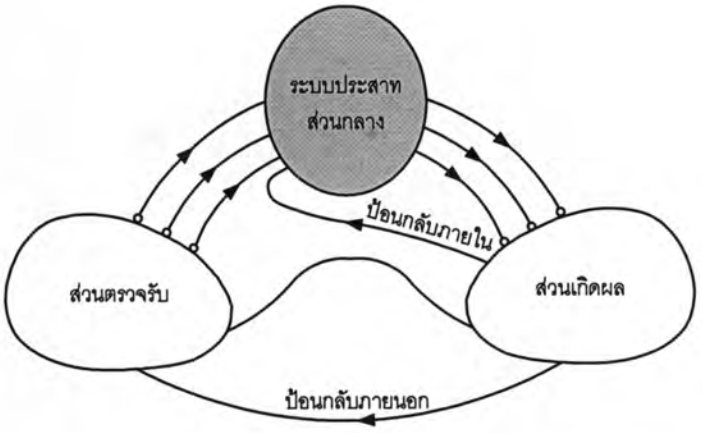
บทที่ 3

ทฤษฎีโครงข่ายประสาทเทียม

เนื้อหาที่จะกล่าวถึงต่อไปในบทนี้ประกอบด้วยหลักการพื้นฐานที่เกี่ยวกับโครงข่ายประสาท โครงสร้างองค์ประกอบของโครงข่ายประสาท กลไกการทำงานของกระบวนการเรียนรู้ทั้งโครงข่ายประสาทธรรมชาติและโครงข่ายประสาทเทียม

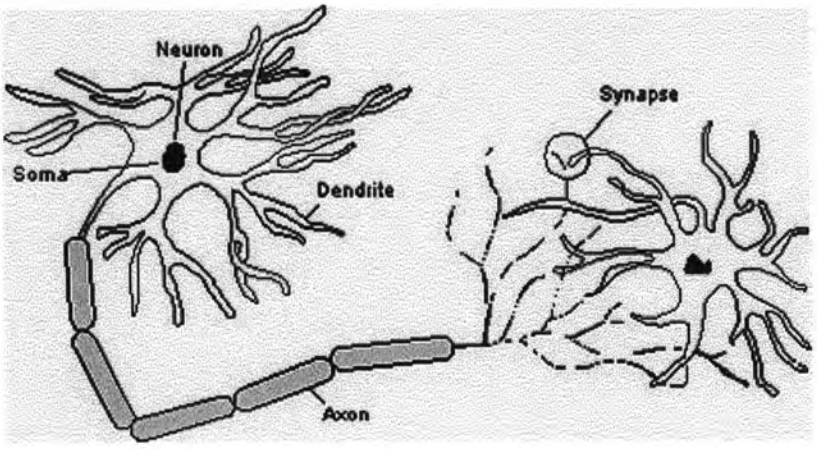
3.1 โครงข่ายประสาทธรรมชาติ

สมองมนุษย์ประกอบด้วยหน่วยการประมวลผล (Computing or processing elements or elementary nerve cell) หรือเรียกว่า นิวรอน (Neuron) ประมาณ10^{11} หน่วย แต่ละนิวรอนเชื่อมโยง (Synapse or Connection) กันเป็นโครงข่าย (Network) จำนวนการเชื่อมโยงมีประมาณ $6x10^{15}$ [4] นั่นคือ มีความหนาแน่นของการเชื่อมโยงประมาณ 10^4 ต่อหนึ่งนิวรอน การทำงานของสมองมนุษย์มีความซับซ้อนไม่เป็นเชิงเส้นและเป็นแบบขนาน (Parallel) การทำงานแต่ละครั้งของสมองมนุษย์จะใช้เวลาอยู่ในช่วงมิลลิวินาที (10^{-3}) [4]-[5] ระบบประสาทของมนุษย์มีการส่งผ่านข้อมูลที่มีการเปลี่ยนรูประหว่างสัญญาณไฟฟ้ากับสัญญาณเคมี เราอาจพิจารณาว่าสมองเป็นเสมือนโครงข่ายการปิดเปิดทางไฟฟ้า (Electrical switching network) ที่ถูกควบคุมด้วยกระบวนการทางชีวเคมี การทำงานของร่างกายมนุษย์ประกอบด้วย 3 ส่วนหลักคือ ส่วนตรวจรับ (Receptors or sensory units) ระบบประสาทส่วนกลาง (Central nerve system) และส่วนเกิดผล (Effector) ส่วนตรวจรับ จะรับข้อมูลขาเข้าทั้งจากภายนอกและภายในระบบ จากนั้นจะส่งข้อมูลในรูปสัญญาณไฟฟ้าแบบอิมพัลส์ (Electrical impulse) เข้าระบบประสาทส่วนกลางต่อไป ผลลัพธ์ที่ได้จากการประมวลของระบบประสาทส่วนกลางจะส่งไปควบคุมการทำงานของส่วนเกิดผล ดังรูปที่ 3.1 สรุปกระบวนการที่เกิดในระบบประสาท คือมีการรับข้อมูล ส่งผ่านข้อมูล ประมวลผลข้อมูล เปรียบเทียบผลที่ได้กับข้อมูลที่มีอยู่ในระบบประสาทส่วนกลาง และอาจทำให้เกิดผลขึ้นมา จะเห็นได้ว่ามีส่วนป้อนกลับ (Feedback) ที่ใช้ในการควบคุมให้เกิดผลตามที่ต้องการ



รูปที่ 3.1 องค์ประกอบของโครงข่ายประสาทของสมองมนุษย์

โครงข่ายประสาทธรรมชาติ ประกอบด้วยองค์ประกอบหลัก 3 ส่วน คือ โซมา (Soma) หรือ ตัวเซล (Cell body) เด็นไดร์ท (Dendrite) และแอกซอน (Axon) ดังรูปที่ 3.2



รูปที่ 3.2 เซลล์ประสาท

เด็นไดรท์ คือส่วนที่มีกิ่งก้านคล้ายต้นไม้ทำหน้าที่เป็นส่วนรับข้อมูลและส่งให้กับโซมา โซมา คือส่วนที่รวมข้อมูลขาเข้าทั้งหมดเข้าด้วยกันและประมวลผล แอกซอน คือส่วนที่มีรูปร่างเป็นทรงกระบอกยาวทำหน้าที่นำข้อมูลออกจากโซมา ในการเชื่อมนิวรอนเข้าด้วยกันเป็นระบบโครงข่ายยังมีองค์ประกอบหนึ่งที่อยู่ระหว่างแอกซอนของนิวรอนหนึ่งกับเดนไดรท์ของนิวรอนที่อยู่ข้างเคียง (Neightboring neural) เรียกว่าไซแนป (Synapse) ไซแนปจะสร้างการกระตุ้นขึ้นมา 2 แบบ คือ แบบเสริม (Exitation) และกระตุ้นหักล้าง (Inhibition) ในทางปฏิบัติจะแทนการกระตุ้น

ทั้งสมองแบบด้วยเครื่องหมายบวกและเครื่องหมายลบตามลำดับ อย่างไรก็ตามนิวรอนจะไม่รับข้อมูลทั้งสองแบบในเวลาเดียวกัน

การส่งผ่านข้อมูลระหว่างนิวรอนโดยปกติจะเป็นผลมาจากตัวโอนถ่าย (Transmitters) แบบเคมี แต่มีบางครั้งที่เป็นผลจากตัวถ่ายโอนแบบไฟฟ้า นิวรอนจะตอบสนองต่อข้อมูลขาเข้าในช่วงเวลาสั้นๆ หากศักย์ดา (Potential) ทั้งหมดที่นิวรอนสูงกว่าขีดจำกัดค่าหนึ่งนิวรอนก็จะสร้างอิมพัลส์ อิมพัลส์ที่ป้อนเข้าสู่นิวรอนจะมาจาก 2 ส่วนเท่านั้น คือจากนิวรอนข้างเคียงและจากตัวมันเอง โดยปกติการกระตุ้นนิวรอนหนึ่งจะใช้อิมพัลส์จำนวนที่คงที่

อิมพัลส์เหล่านี้จะถูกส่งออกมากระตุ้นนิวรอนอื่นๆในเวลาที่พร้อมกัน (Synchronous) รูปแบบของข้อมูลที่ส่งผ่านจะอยู่ในรูปเลขฐานสอง (Binary signal) การกระตุ้นจะมีผลต่อนิวรอน ณ เวลา k หรือไม่ขึ้นอยู่กับการกระตุ้นนิวรอนนั้น ณ เวลา k-1 ก่อนหน้านี้ หากผลต่างระหว่างจำนวนกระตุ้น ณ เวลา k เปรียบเทียบกับจำนวนการกระตุ้น ณ เวลา k-1 เป็นจำนวนมากกว่าค่าขีดจำกัด (Treshold) เราอาจสรุปคุณสมบัติของโครงข่ายประสาทธรรมชาติได้ดังนี้ [6]

1. แต่ละนิวรอนรับข้อมูลขาเข้าเป็นจำนวนมาก
2. ข้อมูลขาเข้าอาจถูกปรับให้เหมาะสมด้วยค่าการเชื่อมโยง (Connection value) หรือ ค่าถ่วงน้ำหนัก (Weight value) ที่ไซแนป
3. โซมาจะรวมข้อมูลขาเข้าที่ผ่านการปรับด้วยค่าถ่วงน้ำหนักด้วยกันและทำการประมวลผล
4. ภายใต้ภาวะที่เหมาะสม เช่น ข้อมูลขาเข้ามีจำนวนเพียงพอ นิวรอนจะส่งข้อมูลออกมาหนึ่งค่า
5. ข้อมูลขาออกจากนิวรอนหนึ่งอาจจะถูกส่งออกไปให้นิวรอนอื่นๆ หลายนิวรอนผ่านทางแอกซอน
6. นิวรอนแต่ละนิวรอนมีความเป็นเฉพาะตำแหน่งหรือเฉพาะที่ (Local)
7. ความจำแบ่งออกเป็น 2 ส่วนคือ
   - ความจำระยะยาว (Long-term) จะเก็บที่ไซแนปของนิวรอนหรือเก็บในค่าถ่วงน้ำหนัก
   - ความจำระยะสั้น (Short-term) สัมพันธ์กับข้อมูลขาออกจากนิวรอน
8. ความถูกต้องของค่าถ่วงน้ำหนักจะขึ้นกับประสบการณ์ในการเรียนรู้ที่ผ่านมา
9. การควบคุมการถ่ายโอนของไซแนปอาจจะเป็นการกระตุ้นแบบเสริมหรือแบบหักล้าง

นอกจากที่กล่าวมาโครงข่ายประสาทธรรมชาติยังยืดหยุ่นต่อความผิดพร่อง (Fault-tolerance) 2 ประเด็นด้วยกัน ประการแรกคือ ความสามารถในการจดจำและจำแนกข้อมูลขาเข้าที่แตกต่างจากข้อมูลขาเข้าที่เคยได้รับไม่มากนัก ตัวอย่างเช่น ความสามารถในการจำแนกภาพบุคคลที่ไม่เคยเห็นมาก่อน หรือสามารถจดจำบุคคลที่ไม่เคยพบเห็นมาก่อน หรือสามารถจดจำบุคคลที่ไม่ได้พบเห็นกันเป็นเวลานาน ประการที่สองความสามารถในการซ่อมแซมหรือชดเชย

ความเสียหายของตัวนิวรอนเอง ในสมองมนุษย์เมื่อนิวรอนหนึ่งชำรุดเสียหายก็จะมีการสร้างนิวรอนบางส่วน ในบางครั้งนิวรอนที่เหลือจะได้รับการปรับสอน (Training) เพื่อทำหน้าที่แทนนิวรอนที่ชำรุดเสียหาย

จากจุดเด่นหลายประการของโครงข่ายประสาทธรรมชาติผนวกกับเทคโนโลยีด้านคอมพิวเตอร์ การทำงานแต่ละครั้งของอุปกรณ์สารกึ่งตัวนำในคอมพิวเตอร์จะใช้เวลาอยู่ในช่วงนาโนวินาที (10^{-9}s) [4]-[5] ซึ่งมีความรวดเร็วมากเมื่อเปรียบเทียบกับการทำงานของสมองมนุษย์ ทำให้มีการคิดค้นแนวทางแก้ปัญหาโดยมีโครงสร้างองค์ประกอบและกลไกการแก้ปัญหาลักษณะเดียวกับโครงข่ายธรรมชาติ โดยเรียกว่า โครงข่ายประสาทเทียม (Artificial neural network, neurocomputers, connectionist networks, parallel distributed processors) [5] หรือเรียกอย่างง่ายว่าโครงข่ายประสาท (Neural network) ในวิทยานิพนธ์ที่จะกล่าวถึงต่อไป โครงข่ายประสาทหมายถึงโครงข่ายประสาทเทียม ยกเว้นเว้นระบุไว้เป็นอย่างอื่น

โครงข่ายประสาทจัดว่าเป็นกลุ่มของตัวประมวลผลแบบขนานที่มีการเก็บความรู้เก็บประสบการณ์และสามารถนำมาใช้ประโยชน์ได้โดยที่

- ความรู้ (Knowledge) จากโครงข่ายประสาทเกิดจากกระบวนการเรียนรู้
- ความรู้จะถูกเก็บไว้ที่ค่าการเชื่อมโยงหรือค่าถ่วงน้ำหนักของนิวรอน

3.2 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมเป็นการใช้ประโยชน์จากความก้าวหน้าของสารกึ่งตัวนำ เป็นระบบประมวลผลข้อมูลที่มีพฤติกรรมบางอย่างเหมือนกับโครงข่ายประสาทธรรมชาติ โครงข่ายประสาทเทียมอาจพิจารณาให้เป็นขั้นตอนวิธีทางคณิตศาสตร์ (Mathematical algorithm) ของโครงข่ายธรรมชาติ โดยตั้งบนพื้นฐานดังนี้

1. หน่วยประมวลผลเรียกว่า นิวรอน การส่งผ่านข้อมูลระหว่างนิวรอนทำได้โดยการเชื่อมโยงระหว่างนิวรอน
2. การเชื่อมโยงแต่ละส่วนจะขึ้นอยู่กับค่าถ่วงน้ำหนัก
3. ข้อมูลขาออกถูกกำหนดด้วยฟังก์ชันกระตุ้น (Activation function หรือ Transfer function) ความสัมพันธ์ระหว่างผลรวมของข้อมูลขาเข้าที่คูณด้วยค่าถ่วงน้ำหนักข้อมูลขาออกเป็นความสัมพันธ์แบบไม่เป็นเชิงเส้น (Non-linear)

คุณสมบัติของโครงข่ายประสาทเทียมจึงถูกกำหนดด้วยองค์ประกอบทั้งสามส่วน นั้นคือ ส่วนแรกในข้อ 1 เป็นรูปแบบการเชื่อมโยงระหว่างนิวรอนเรียกส่วนนี้ว่า สถาปัตยกรรมของ

โครงข่าย (Network architecture) ส่วนต่อไปในข้อ 2 เป็นการกำหนดค่าถ่วงน้ำหนัก เรียกส่วนนี้ว่า ขั้นตอนวิธีการฝึกสอนหรือการเรียนรู้ ส่วนสุดท้ายคือข้อ 3 เกี่ยวกับการกำหนดข้อมูลขาออกโดยฟังก์ชันกระตุ้น

โครงข่ายประสาทเทียมได้รับการพัฒนามาเป็นเวลาประมาณ 54 ปี [6]-[7] ตั้งแต่ ค.ศ. 1943 โดย McCulloch-Pitts โครงข่ายประสาทเทียมพัฒนาขึ้นมาเพื่อเพิ่มศักยภาพในการคำนวณให้มีประสิทธิภาพมากขึ้น โดยกำหนดค่าถ่วงน้ำหนักที่ทำให้นิวรอนสามารถจัดการกับฟังก์ชันตรรก (Logic function) โดยนิวรอนรูปแบบหนึ่งจะใช้จัดการกับฟังก์ชันตรรกหนึ่งๆ รูปแบบที่พัฒนาขึ้นโดย McCulloch-Pitts ใช้ได้เป็นอย่างดีกับวงจรทางตรรก ต่อมาได้มีการพัฒนาโครงข่ายประสาทเทียมทั้งแบบเดียวกัน McCulloch-Pitts และแนวทางอื่นๆอีกมากมาย ตัวอย่างเช่น ปี ค.ศ. 1949 Hebb ได้เสนอกฎการเรียนรู้ (Learning rule) ซึ่งเป็นที่รู้จักกันในนามกฎการเรียนรู้แบบเฮบเบียน (Hebbian) ปีค.ศ. 1958 Rosenblatt ได้พัฒนาโครงข่ายประสาทที่เรียกว่าเพอร์เซ็ปตรอน (Perceptrons) ปีค.ศ. 1943 Widrow และ Hoff ได้พัฒนาโครงข่ายประสาท ADALINE (ADAtive LINEar combiner) และต่อมาได้พัฒนามาเป็น MADALINE (Many ADALIN) ปีค.ศ. 1972 Kohonen ได้พัฒนา SOFM (Self-Oranizing feature maps) ในปีค.ศ. 1967 Grossberg เริ่มพิมพ์ผลงานของเขาจนเป็นที่รู้จักอย่างกว้าขวางและปี ค.ศ 1985 Grossberg และ Carpenter ได้ร่วมกันพัฒนาโครงข่ายทฤษฎีอะแดปตีฟ เรโนแนนท์ (Adaptive resonance theory nets) สำหรับข้อมูลขาเข้าแบบฐานสองในนามของ ART1 และ ART2 ในเวลาต่อมา ในปี ค.ศ. 1986 Rumelhart, Hilton และ Willams และปีค.ศ. 1988 McClelland และ Rumellhart ได้พัฒนาโครงข่ายประสาทซึ่งเป็นที่รู้จักกันดีในนามการแพร่กระจายกลับ (Back-Propagation) ในปีค.ศ.1982 และต่อมาปี ค.ศ. 1985 Hopfield ส่งผลให้ Hopfield ได้รับราวัลโนเบล สาขาวิชา ฟิสิกส์ ช่วงปี ค.ศ. 1975 ถึง ค.ศ.1988 Fukushime และคณะได้พัฒนาโครงข่ายประสาทสำหรับ จดจำและจำแนกอักษรในนามของ Neurocognitrol ช่วงปีค.ศ. 1983 ถึง ค.ศ. 1987 นักวิจัยหลายท่านได้พัฒนาการหาค่าถ่วงน้ำหนักและฟังก์ชันกระตุ้น โดยใช้ฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability density function) ต่างจากเดิมที่มักจะใช้แบบนอนดีเทอร์มินิสติก (Nondeterministic) โดยพัฒนาโครงข่ายประสาทในนาท โบลท์ มานน์ (Boltzmann) อาศัยแนวคิดการจำลองการอ่อนตัวลงสู่ภาวะเสถียรภาพ (Simulated annealing) และทฤษฎีการตัดสินใจของบาเยเซียน (Bayesian decision theory) นอกจากที่กล่าวมายังได้พัฒนาในรูปแบบของฮาร์ดแวร์ในรูปแบบต่างๆ เช่น VLSI ไมโครชิพ เป็นต้น

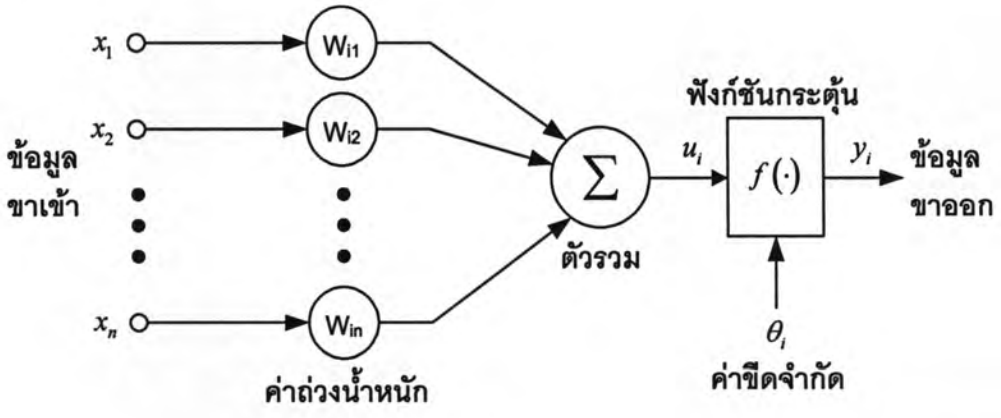
นับตั้งแต่การตีพิมพ์ผลงานของ Hopfield ในปีค.ศ. 1982 และหนังสือที่เขียนทั้งสองเล่มโดย Rumelhart และ McClelland ในปี ค.ศ. 1986 ส่งผลทำให้เกิดความสนใจเกี่ยวกับโครงข่ายประสานกันอย่างกว้าง ทำให้โครงข่ายประสาทได้รับการพัฒนาในด้านต่างๆ เช่น วิทยาศาสตร์

ด้านระบบประสาท (Neurosciences) วิทยาศาสตร์กายภาพ (Physical sciences) การศึกษาเกี่ยวกับจิตวิทยา ด้านคณิตศาสตร์ และด้านวิศวกรรมเป็นต้น สามารถกล่าวได้ว่า โครงข่ายประสาทยังคงได้รับการพัฒนาทั้งด้านทฤษฎี การออกแบบและการประยุกต์ใช้อีกต่อไปในอนาคต

แนวทางที่จะกล่าวถึงโครงข่ายประสาทแบ่งออกได้เป็น 3 ส่วนด้วยกัน คือ แบบจำลองของโครงข่ายประสาท โครงสร้างสถาปัตยกรรมของโครงข่ายประสาทหรือการเชื่อมโยงนิวรอน และการเรียนรู้ในรูปแบบต่างๆ

3.2.1 แบบจำลองของโครงข่ายประสาท

นิวรอน เป็นหน่วยประมวลผลพื้นฐานที่สำคัญในการทำงานของโครงข่ายประสาท มีองค์ประกอบ 3 ส่วน รูปที่ 3.3 คือ



รูปที่ 3.3 การจำลองโครงข่ายประสาทในรูปแบบคณิตศาสตร์

1. กลุ่มของการเชื่อมโยงหรือไซแนป การเชื่อมโยงระหว่างนิวรอนจะถูกควบคุมด้วยค่าถ่วงน้ำหนัก กำหนดให้ x_j เป็นข้อมูลขาเข้าของไซแนป j ที่ป้อนให้กับนิวรอน i ข้อมูลขาเข้าจะถูกคูณด้วยค่าถ่วงน้ำหนัก w_{ij} โดยที่ตัวห้อยตัวแรก (i) หมายถึงนิวรอนที่ i ส่วนตัวที่สอง (j) บอกว่าเป็นข้อมูลขาเข้าที่ j ซึ่งอาจมาจากนิวรอนอื่นๆ ที่อยู่ข้างเคียง หาก w_{ij} มีเครื่องหมายเป็นบวกหมายถึงการกระตุ้นเสริม ถ้ามีเครื่องหมายเป็นลบหมายถึงการกระตุ้นแบบหักล้าง
2. ตัวบวก (Summation หรือ Adder) สำหรับรวมข้อมูลขาเข้าที่ถูกคูณด้วยค่าถ่วงน้ำหนัก เราเรียกว่า การรวมเชิงเส้น (Linear combiner)
3. ฟังก์ชันกระตุ้น เพื่อจำกัดช่วงข้อมูลขาออกให้อยู่ในช่วงต้องการ

จาก รูปที่ 3.3 สามารถอธิบายนิวรอน i อยู่ในรูปสมการได้ด้วยสมการ 2 สมการ คือ

$$u_i = \sum_{j=1}^{n} w_{ij} x_j \tag{3.1}$$

$$y_i = f\left(u_i + \theta_i\right) \tag{3.2}$$

โดยที่

$x_1, x_2, \ldots, x_n$ คือ ข้อมูลขาเข้า

$w_{i1}, w_{i2}, \ldots, w_{in}$ คือ ค่าถ่วงน้ำหนักของนิวรอน i

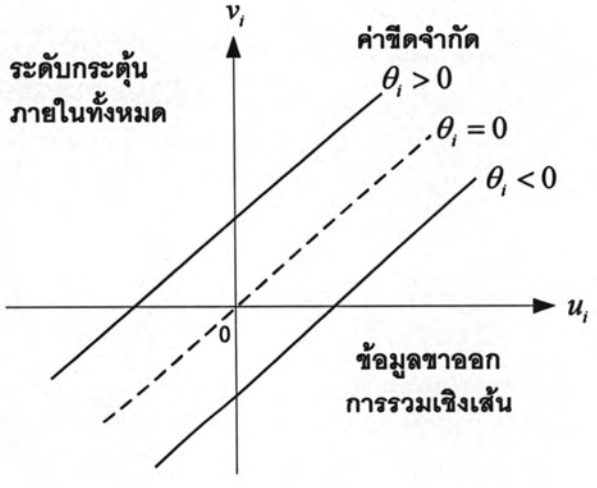
u_i คือ ข้อมูลขาออกในนิวรอน (Internal output) หรือ ข้อมูลขาเข้าสุทธิ (net_i)

θ_i คือ ค่าขีดจำกัด (Threshold) หรือ ไบแอส (Bias) ซึ่งรับมาจากภายนอก

$f(\cdot)$ คือ ฟังก์ชันกระตุ้น

y_i คือ ข้อมูลขาออก

การป้อนค่าขีดจำกัดให้กับนิวรอน จะได้ความสัมพันธ์ระหว่างระดับกระตุ้นภายใน (Internal activity level หรือ Internal activation potential) ค่าขีดจำกัด และข้อมูลขาเข้า ดังต่อไปนี้



รูปที่ 3.4 ผลของค่าขีดจำกัด

$$v_i = u_i + \theta_i \tag{3.3}$$

โดยที่

v_i คือ ระดับกระตุ้นภายใน

จากโครงข่ายประสาทธรรมชาติและแบบจำลองของโครงข่ายประสาท ซึ่งแบบจำลองทางคณิตศาสตร์ที่ประดิษฐ์ขึ้นมา สามารถเปรียบเทียบได้ดังตารางที่ 3.1

ตารางที่ 3.1 เปรียบเทียบระหว่างนิวรอนธรรมชาติและแบบจำลองทางคณิตศาสตร์ที่ประดิษฐ์ขึ้น

ลำดับ	โครงข่ายประสาทธรรมชาติ	แบบจำลองโครงข่ายประสาทที่ประดิษฐ์ขึ้น
1	โซมา	ส่วนประมวลผล(ตัวรวมและฟังก์ชันกระตุ้น)
2	เดนไดรท์	ข้อมูลขาเข้า
3	แอกซอน	ข้อมูลขาออก
4	ไซแนป	ค่าถ่วงน้ำหนัก
5	ความเร็วในการประมวลผลต่ำ	ความเร็วในการประมวลผลสูง
6	โครงข่ายประกอบด้วยนิวรอนจำนวนมาก	โครงข่ายประกอบด้วยนิวรอนจำนวนน้อย

3.2.2 ชนิดของฟังก์ชันกระตุ้น

ฟังก์ชันกระตุ้นที่แทนด้วย $f(\cdot)$ เป็นตัวกำหนดค่าข้อมูลขาออก ฟังก์ชันส่วนขาเข้าที่นิยมใช้ คือ

ก. ฟังก์ชันเชิงเส้น (Linear function)

$$net_i = \sum_{i=1}^{n} w_{ij} x_j + \theta_i \tag{3.4}$$

ข. ฟังก์ชันกำลังสอง (Quadratic function)

$$net_i = \sum_{i=1}^{n} w_{ij} (x_j)^2 + \theta_i \tag{3.5}$$

เมื่อ

net_i คือ ข้อมูลขาเข้าสุทธิ

x_j คือ ข้อมูลขาเข้านิวรอนที่ j

w_{ij} คือ ค่าคงที่ถ่วงน้ำหนักที่เชื่อมโยงระหว่างนิวรอน j ในชั้นข้อมูลขาเข้า และ นิวรอน i ในชั้นซ่อน

θ_i คือ ค่าขีดจำกัด หรือ ไบแอส

ส่วนฟังก์ชันส่วนขาออกมีฟังก์ชันที่นิยมใช้ คือ

ก. ฟังก์ชันเชิงเส้น (Linear function)

$$f(net_i) = net_i \tag{3.6}$$

ข. ฟังก์ชันขั้นบันได (Step function)

$$f(net_i) = \begin{cases} 1; net_i \geq 0 \\ 0; net_i < 0 \end{cases} \tag{3.7}$$

ค. ฟังก์ชันล็อกซิกมอยด์ (Log-Sigmoid function)

$$f(net_i) = \frac{1}{1+e^{-net_i}} \tag{3.8}$$

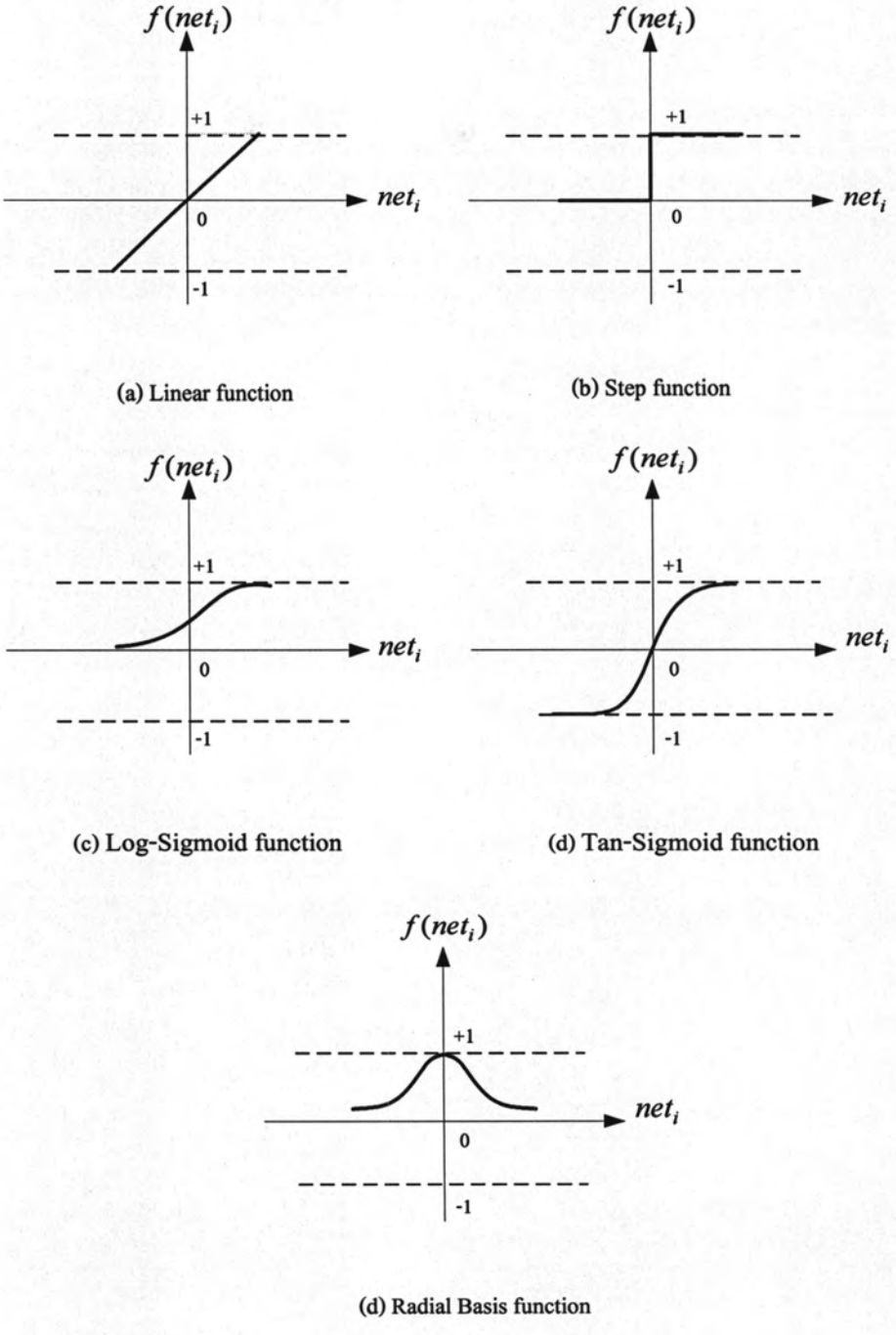
ง. ฟังก์ชันแทนซิกมอยด์ (Tan-Sigmoid function)

$$f(net_i) = \frac{2}{\left(1+e^{(-2*net_i)}\right)} - 1 \tag{3.9}$$

จ. ฟังก์ชันเรเดียลเบซีส (Radial Basis function)

$$f(net_i) = \frac{1}{e^{(net_i)^2}} \tag{3.10}$$

จาก (3.6)-(3.10) สามารถเขียนรูปกราฟความสัมพันธ์ระหว่างข้อมูลขาเข้าสุทธิกับข้อมูลขาออกได้ดังรูปที่ 3.5



(a) Linear function

(b) Step function

(c) Log-Sigmoid function

(d) Tan-Sigmoid function

(d) Radial Basis function

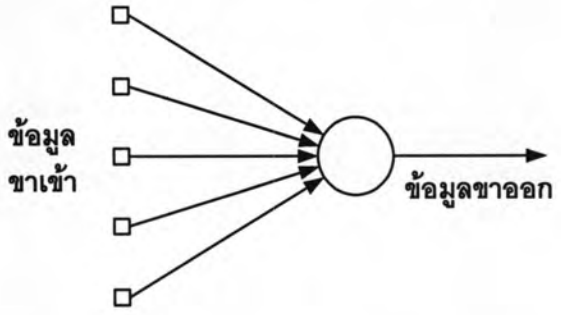
รูปที่ 3.5 ความสัมพันธ์ระหว่างข้อมูลขาเข้าสุทธิกับข้อมูลขาออก

สำหรับการพิจารณาเลือกใช้ฟังก์ชันกระตุ้นแบบใดนั้นนับถึงปัจจุบันยังไม่มีข้อกำหนดที่แน่นอน แต่อาจจะพิจารณาจากองค์ประกอบดังต่อไปนี้

1. ความเป็นเชิงเส้น หรือไม่เป็นเชิงเส้น
2. ลักษณะของข้อมูลขาเข้า
3. ขอบเขตที่ต้องการและลักษณะของข้อมูลขาออก
4. ปัจจัยอื่นๆที่มีนัยสำคัญต่อผลตอบสนองของนิวรอนของโครงข่ายประสาทโดยรวม

3.3 สถาปัตยกรรมของโครงข่ายประสาท (Neural network architecture)

เมื่อพิจารณาถึงการเชื่อมโยงของนิวรอนกล่าวได้ว่าโครงข่ายประสาทนอกจากจะเป็นวิทยาศาสตร์แล้วยังเป็นศิลปศาสตร์ด้วย การออกแบบด้านสถาปัตยกรรมหรือโครงสร้างของโครงข่ายประสาทนี้จะเกี่ยวกับเส้นทางการเชื่อมโยง ทิศทางการส่งผ่านข้อมูลในเส้นทาง จำนวนชั้นทั้งหมด จำนวนนิวรอนทั้งหมดที่ประกอบเป็นโครงข่ายประสาท และจำนวนนิวรอนต่อชั้น เป็นต้น เพื่อให้โครงสร้างของโครงข่ายประสาทดูง่ายขึ้นจึงใช้แบบจำลองอย่างย่อดังรูปที่ 3.6 แทนแบบจำลองใน รูปที่ 3.3

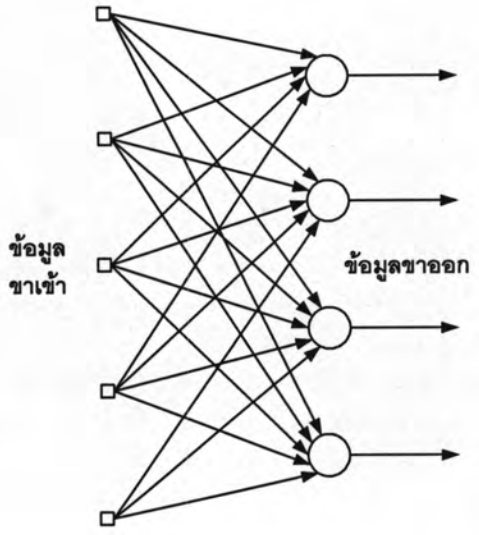


รูปที่ 3.6 แบบจำลองโครงข่ายประสาทอย่างง่าย

จากภาพแบบจำลองโครงข่ายประสาทขนาดย่อจะเห็นว่ารวมเอาส่วนฟังก์ชันกระตุ้นเข้ากับส่วนตัวบวกซึ่งเรียกว่าส่วนประมวลผล (Processing part หรือ Computation part) และในแบบจำลองขนาดย่อนี้ไม่แสดงค่าเทรสโฮลด์

3.3.1 โครงข่ายป้อนไปข้างหน้าชั้นเดียว (Single- layer feedforword networks)

โครงข่ายป้อนไปข้างหน้าชั้นเดียว จัดเป็นรูปแบบของโครงสร้างโครงข่ายประสาทที่ซับซ้อนน้อยที่สุด ชั้นของโครงข่ายประกอบด้วยกลุ่มของนิวรอน การส่งผ่านข้อมูลกระทำจากกลุ่มข้อมูลขาเข้าส่งไปยังชั้นข้อมูลขาออก และไม่มีการส่งข้อมูลย้อนกลับ จึงเรียกรูปแบบการส่งผ่านข้อมูลแบบนี้ว่า ป้อนไปข้างหน้า (Feedforward) โครงข่ายที่แสดงใน รูปที่ 3.7 เรียกว่า โครงข่ายชั้นเดียว (Single-layer network) การนับจำนวนชั้นของโครงข่ายในปัจจุบันจะนับเป็นชั้นเมื่อชั้นนั้นมีโหนดที่เป็นส่วนประมวลผล กล่าวคือ ไม่นับกลุ่มข้อมูลขาเข้าภายนอกของโครงข่ายเป็นชั้นเนื่องจากไม่มีส่วนประมวลผลอยู่นั่นเอง

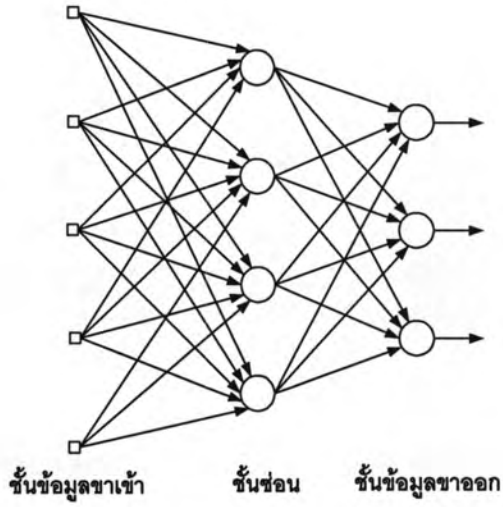


รูปที่ 3.7 โครงข่ายป้อนไปข้างหน้าชั้นเดียว

3.3.2 โครงข่ายป้อนไปข้างหน้าหลายชั้น (Multilayer feedorward networks)

โครงสร้างของโครงข่ายแบบที่สองเป็นแบบป้อนไปข้างหน้าเช่นกัน แต่แตกต่างจากแบบแรก เนื่องจากมีชั้นซ่อนอย่างน้อยหนึ่งชั้น ชั้นซ่อนหรือนิวรอนที่ซ่อน (Hidden neuron) จะมีส่วนประมวลผลอยู่ด้วย นิวรอนที่ซ่อนอยู่นี้จะแทรกอยู่ระหว่างส่วนข้อมูลขาเข้าจากภายนอกของโครงข่ายและส่วนข้อมูลขาออกของโครงข่าย ข้อดีประการหนึ่ง ของการมีชั้นเพิ่มเข้ามานี้คือมีความสามารถทางสถิติในอันดับสูง (High– order statistics)ขึ้น ซึ่งเป็นประโยชน์อย่างชัดเจนเมื่อข้อมูลขาเข้ามีจำนวนมาก

ข้อมูลขาเข้าโหนดต้นทางภายในกลุ่มข้อมูลขาเข้าของโครงข่ายจะถูกป้อนให้ส่วนประมวลผลของชั้นต่อไปหรือชั้นซ่อนชั้นแรกนั่นเอง ข้อมูลขาออกที่ได้จากชั้นซ่อนชั้นแรกนี้จะถูกส่งให้กับชั้นที่อยู่ต่อไป ในฐานะข้อมูลขาเข้าของชั้นที่รับข้อมูลนั้น หากยังมีชั้นซ่อนเหลืออยู่ก็จะรับข้อมูลที่ส่งออกจากชั้นที่อยู่ก่อนหน้านี้ สำหรับข้อมูลขาออกของโครงข่ายได้จากชั้นข้อมูลขาออก ตัวอย่างของโครงข่ายป้อนไปข้างหน้าหลายชั้นดังรูปที่ 3.8 เรียกว่าโครงข่ายแบบ 5-4-3 โหนดต้นทางของกลุ่มข้อมูลขาเข้าจำนวน 5 โหนด มีจำนวนนิวรอนชั้นซ่อน 4 นิวรอน และมีนิวรอนข้อมูลขาออกจำนวน 3 นิวรอน ในกรณีที่โครงข่ายมีโหนดต้นทางของกลุ่มข้อมูลขาเข้าจำนวน p โหนด จำนวนนิวรอนในชั้นซ่อนชั้นแรก h_1 นิวรอนในชั้นที่สอง h_2 นิวรอน และจำนวนนิวรอนในชั้นข้อมูลขาออกจำนวน q นิวรอน เราเรียกโครงข่ายแบบนี้ว่าโครงข่าย p-h_1-h_2-q

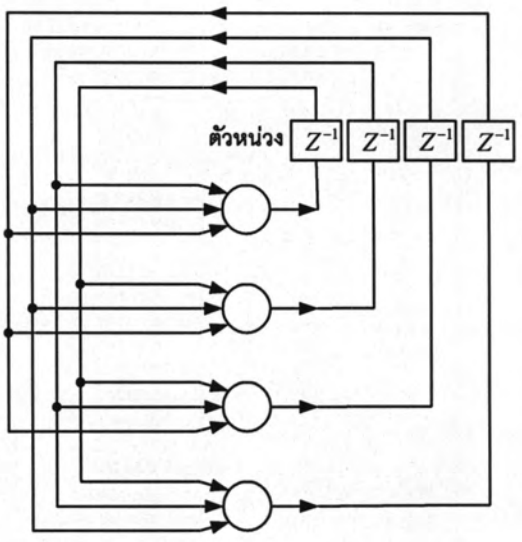


รูปที่ 3.8 โครงข่ายป้อนไปข้างหน้าหลายชั้นแบบเชื่อมโยงอย่างเต็มที่

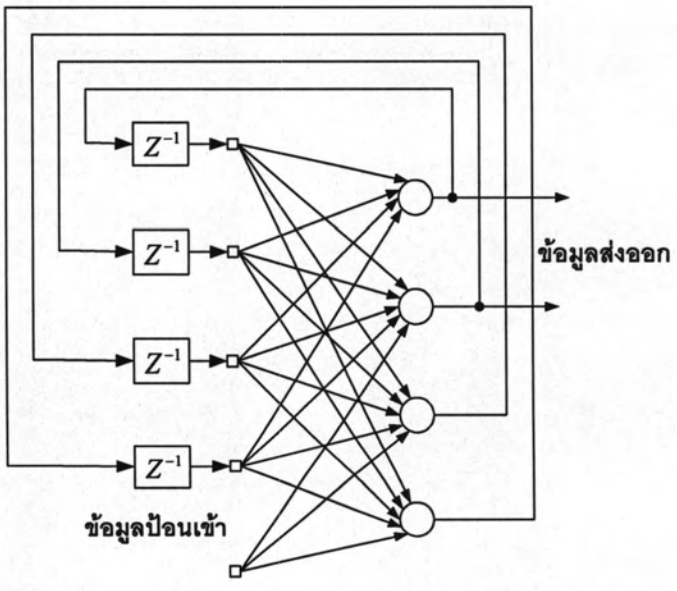
จากรูปที่ 3.8 เป็นการเชื่อมโยงที่เรียกว่า เชื่อมโยงอย่างเต็มที่ (Full connected) การเชื่อมโยงอย่างเต็มที่ในกรณีที่โครงข่ายเป็นแบบป้อนไปข้างหน้านี้หมายความว่า แต่ละโหนดหรือแต่ละนิวรอนของชั้นที่อยู่ก่อนหน้านี้จะมีเส้นทางเชื่อมโยงไปยังทุกๆ นิวรอนที่อยู่ในชั้นต่อไป

3.3.3 โครงข่ายย้อนกลับ (Recurrent networks)

โครงข่ายย้อนกลับนี้ต่างจากโครงข่ายป้อนไปข้างหน้า 2 แบบแรกที่กล่าวมา ตรงที่มีวงรอบป้อนกลับ (Feed back loop) เราเรียกโครงข่ายแบบนี้ว่า โครงข่ายป้อนกลับ จากรูปที่ 3.9 เป็นโครงข่ายป้อนกลับที่ไม่มีชั้นซ่อน ซึ่งอาจเรียกได้ว่าเป็นโครงข่ายย้อนกลับชั้นเดียว แต่ละนิวรอนจะส่งข้อมูลขาออกของมันป้อนกลับไปเป็นข้อมูลขาเข้าของนิวรอนอื่นๆ และจะไม่มีวงรอบใดที่ป้อนกลับให้ตัวมันเอง (Self-Feedback) สำหรับรูปที่ 3.10 เป็นตัวอย่างของโครงข่ายป้อนกลับหลายชั้นประกอบด้วยชั้นซ่อนหนึ่งชั้น จากรูปโครงข่ายย้อนกลับทั้งสองจะพบว่ามีส่วนประกอบเพิ่มเข้ามาคือ ตัวหน่วง (Unit- delay elements) แทนด้วยสัญลักษณ์ Z^{-1} ทั้งนี้เพื่อให้โครงข่ายสามารถแสดงพฤติกรรมภาวะไดนามิกส์ไม่เป็นเชิงเส้น (Nonlinear dynamics) ซึ่งเป็นภาวะที่สำคัญของโครงข่าย



รูปที่ 3.9 โครงข่ายย้อนกลับชั้นเดียว



รูปที่ 3.10 โครงข่ายย้อนกลับชั้นหลายชั้น

สถาปัตยกรรมของโครงข่ายประสาทที่กล่าวมาเป็นสถาปัตยกรรมพื้นฐานที่ไม่ซับซ้อนมาก นักในงานจริงที่ประยุกต์ใช้โครงข่ายประสาทในการแก้ปัญหาอาจประกอบด้วยโครงสร้างโครงข่ายประสาทพื้นฐานมากกว่า 1 แบบทั้งนี้เพื่อให้มีความสามารถเพียงพอในการแก้ปัญหานั่นเอง

3.4 กระบวนการเรียนรู้และกฎการเรียนรู้ของโครงข่ายประสาท

จากหัวข้อที่กล่าวมาทั้งสองส่วน ยังไม่เพียงพอที่จะใช้ประโยชน์จากโครงข่ายประสาท การที่จะทำให้โครงข่ายประสาทมีความสามารถในการหาได้ ต้องฝึกสอน (Training) โครงข่ายประสาท หรือกล่าวอีกอย่างก็คือ โครงข่ายประสาทต้องมีความสามารถที่จะเรียนรู้ กระบวนการฝึกสอนหรือกระบวนการเรียนรู้จะเกี่ยวกับความรู้ ความรู้ที่กล่าวถึงนี้คือข้อมูลข่าวสารแบบจำลองหรือสิ่งอื่นใดที่คนหรือเครื่องจักรเก็บไว้ใช้ในกระบวนการพยากรณ์ ทำนาย ประมาณการ คาดคะเน และตอบสนองอย่างเหมาะสมต่อสถานการณ์หรือสิ่งที่กระตุ้นภายนอก นั่นคือสิ่งที่เกี่ยวกับความรู้จะมี 2 ประเด็น ประเด็นที่หนึ่งเกี่ยวกับความรู้ว่าเป็นข่าวสารข้อมูลอะไร ประเด็นที่สองคือ วิธีการเก็บข่าวสารข้อมูลเพื่อนำมาใช้ประโยชน์นั้นทำอย่างไร เนื่องจากผลลัพธ์ที่ดีจะขึ้นอยู่กับความถูกต้องแม่นยำของตัวของผลลัพธ์เองและความรวดเร็วที่จะได้ผลลัพธ์นั้นมา โครงข่ายประสาทที่ดีจึงควรเป็นแบบจำลองภาวะในการเก็บความรู้พร้อมทั้งสามารถประมวลผลได้ด้วย การเรียนรู้ของโครงข่ายประสาทเป็นกระบวนการทำให้เกิดความรู้ สามารถกล่าวได้ว่าการเรียนรู้ของโครงข่ายประสาทเป็นประบวนการปรับค่าตัวแปรอิสระ (Free parameters) ของโครงข่ายประสาทเพื่อปรับพฤติกรรมให้สอดคล้องกับการกระตุ้นจากสิ่งแวดล้อม กระบวนการเรียนรู้ของโครงข่ายประสาทจึงเป็นกระบวนการทำซ้ำ (Iterative) เราพอจะลำดับกระบวนการเรียนรู้ได้ดังนี้

1. โครงข่ายประสาทถูกกระตุ้นโดยสิ่งแวดล้อม
2. โครงข่ายจะปรับตัวเพื่อตอบสนองต่อสิ่งที่มากระตุ้น
3. โครงข่ายจะตอบสนองต่อสิ่งแวดล้อมด้วยภาวะใหม่หลังจากการปรับตัว

เราอาจจะกล่าวถึงกระบวนการเรียนรู้ใน 2 แนวทาง คือ กฎการเรียนรู้ (Learning rules หรือ Learning algorithm) และรูปแบบการเรียนรู้ (Learning pattern หรือ Learning paradigm)

3.4.1 กฎการเรียนรู้

กฎพื้นฐานในการเรียนรู้ของโครงข่ายประสาท [7] จะเกี่ยวกับการเปลี่ยนแปลงเวคเตอร์ค่าถ่วงน้ำหนัก (Weight vector) $w_i = [w_{i1}\ w_{i2}\ \dots\ w_{in}]^T$ โดยที่ $[]^T$ คือการทรานโพส (Transposition) การเปลี่ยนแปลงนั้นจะขึ้นกับข้อมูลขาเข้า x และสัญญาณการเรียนรู้ (Learning signal) l โดยปกติแล้ว l เป็นฟังก์ชันของ w_i และ c แต่ในบางครั้งอาจมีตัวแปรเพิ่มเติม คือสัญญาณควบคุมการเรียนรู้ (Teacher's signal) c_i นั่นคือ

สัญญาณการเรียนรู้ (l)

$$l = f\left(w_i, x, c_i\right) \tag{3.11}$$

จากกฎพื้นฐานที่กล่าวมาคือ การเปลี่ยนแปลงของเวคเตอร์ของค่าถ่วงน้ำหนัก ณ เวลา t มีความสัมพันธ์ดังนี้

$$\Delta w_i(t) = \alpha l(t) x(t) = \alpha f\left[w_i(t), x(t), c_i(t)\right] x(t) \qquad (3.12)$$

โดย

α คือ ค่าคงที่การเรียนรู้ (Learning constant)

ค่าคงที่การเรียนรู้เป็นตัวกำหนดอัตราการเรียนรู้ (Rate of learning) ให้เหมาะสมไม่เร็วหรือช้าเกินไปค่าดังกล่าวนี้จึงมีค่าเป็นบวก (Positive)

เวกเตอร์ของค่าถ่วงน้ำหนัก ณ เวลา t+1 เป็นผลจากการเปลี่ยนแปลงค่าถ่วงน้ำหนักขั้นตอนการปรับตัวสู่ค่าถ่วงน้ำหนักใหม่นี้เรียกขั้นตอนนี้ว่า ขั้นตอนการเรียนรู้ (Learning step)

$$w_i(t+1) = w_i(t) + \Delta w_i(t) \qquad (3.13)$$

เมื่อนำค่า $w_i(t)$ จาก (3.13) มาแทนจะได้

$$w_i(t+1) = w_i(t) + \alpha f\left[w_i(t), x(t), c_i(t)\right] x(t) \qquad (3.14)$$

เมื่อพิจารณาการปรับค่าเวคเตอร์ของค่าถ่วงน้ำหนักเป็นกระบวนการทำซ้ำ (Iteration) จาก (3.14) สามารถเขียนได้ดังนี้

$$w_i^{k+1} = w_i^k + \alpha f\left[w_i^k, x^k, c_i^k\right] x^k \qquad (3.15)$$

สำหรับตัวยก (Superscript) k+1 หมายถึง การคำนวณในปัจจุบัน ส่วนตัวยก k คือการคำนวณในครั้งที่ผ่านมาก่อนหน้านี้ ในการปรับค่าเวกเตอร์ของค่าถ่วงน้ำหนักนั้น อาจจะเริ่มต้นด้วยการสุ่มเดา (Random) ค่าถ่วงน้ำหนักที่เหมาะสม ตัวอย่างกฎการเรียนรู้ที่มีชื่อเสียงได้แก่

1. กฎการเรียนรู้แบบแฮบเบียน (Hebbian learning rule)
2. กฎการเรียนรู้แบบปรับค่าผิดพลาด (Error correction learning rule)
3. กฎการเรียนรู้แบบคอมแพ็ตทิทีฟ (Competitive learning rule)
4. กฎการเรียนรู้แบบโบลท์มานน์ Bolztmann learning rule)

1. กฎการเรียนรู้แบบแฮบเบียน

การเรียนรู้ของ Hebb นับว่าเป็นกฎการเรียนรู้ที่เก่าแก่และมีชื่อเสียง กล่าวไว้ในหนังสือ Organization of Behavior หน้า 62 ซึ่งตีพิมพ์ในปี ค.ศ. 1949 ดังนี้

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it , some growth process or metabolic changes take place in one or both cells such that A's efficiency as one of the cells firing B, is increased.

จากคำกล่าวข้างบนจะได้สัญญาณการเรียนรู้เป็นฟังก์ชันของข้อมูลขาออกของนิวรอน ดังนี้

$$l \triangleq f\left(w_i^T x\right) \tag{3.16}$$

เวกเตอร์ค่าถ่วงน้ำหนักเปลี่ยนแปลงดังนี้

$$\Delta w_i = \alpha f\left(w_i^T x\right) x \tag{3.17}$$

การเปลี่ยนแปลงค่าถ่วงน้ำหนักแต่ละค่าหาได้จากความสัมพันธ์

$$\Delta w_{ij} = \alpha f\left(w_i^T x\right) x_j \quad ; j = 1, 2, ..., n \tag{3.18}$$

จากกฎการเรียนรู้ของแฮบเบียนจะเห็นว่ารูปแบบการเรียนรู้เป็นแบบไม่มีการควบคุม (Unsupervised) และเป็นแบบป้อนไปข้างหน้า เราสามารถเขียนในรูปย่อได้ดังนี้

$$\Delta w_{ij} = \alpha y_i x_j \tag{3.19}$$

ค่าดังกล่าวเป็นผลคูณระหว่างข้อมูลขาเข้าและข้อมูลขาออก เราอาจจะเรียกว่า Activity product rule ต่อมาในปี ค.ศ. 1988 Kohonen ได้เพิ่มพจน์ที่เป็นฟังก์ชันของ w_{ij} เองเพื่อให้การปรับตัวของค่าถ่วงน้ำหนักเป็นอย่างค่อยเป็นค่อยไป สามารถเขียนสมการได้ดังนี้

$$\Delta w_{ij} = \alpha y_i x_j - \beta y_i w_{ij} \quad ; j = 1, 2, ..., n \tag{3.20}$$

โดย

β คือ ค่าคงที่การเรียนรู้ค่าหนึ่ง

จาก (3.16) สามารถจัดรูปแบบได้ใหม่ดังนี้

$$\Delta w_{ij} = \alpha y_j \left[\gamma x_j - w_{ij} \right] \quad ; \gamma = \frac{\beta}{\alpha} \tag{3.21}$$

จาก (3.20) เรียกว่า กฎการเรียนรู้แบบ Generalized activity product

ในกรณีที่ทราบข้อมูลขาออกที่ต้องการ d_i คือมีสัญญาณควบคุมการเรียน ทำให้การเรียนรู้กลายเป็นการเรียนรู้แบบมีการควบคุม (Supervised) เรียกการเรียนรู้แบบนี้ว่า Correlation learning rule ซึ่งมีความสัมพันธ์ดังนี้

$$\Delta w_{ij} = \alpha d_i x_j \tag{3.22}$$

เมื่อพิจารณา (3.19) และ (3.22) จะเห็นได้ว่าแตกต่างกันที่สัญญาณการเรียนรู้ l ใน (3.19) ใช้ข้อมูลขาออกที่ได้จากนิวรอนจริงๆ y_i ขณะที่ (3.22) ใช้ข้อมูลขาออกที่ต้องการจะได้ d_i สำหรับรูปแบบใน (3.22) เราสามารถเรียกได้ว่าเป็นการเรียนรู้แบบแฮบเบียนที่มีการควบคุม (Supervised Hebbian learning)

นอกจากจำลองทางคณิตศาสตร์ของกฎการเรียนรู้แบบ Hebb ดังที่กล่าวมาได้มีการนำเสนอกฎการเรียนรู้ของ Hebb ในแบบที่ปัญหาเชิงสถิติ การเปลี่ยนแปลงของค่าถ่วงน้ำหนักขึ้นกับค่าโควาเรียนซ์ (Covariance) ระหว่างข้อมูลขาเข้าและข้อมูลขาออก เขียนเป็นความสัมพันธ์ได้ดังนี้

$$\begin{aligned} \Delta w_{ij} &= \alpha \operatorname{cov}[y_i, x_i] \, ; j = 1, 2, ..., n \\ &= \alpha E[(y_i - \overline{y}_i)(x_i - \overline{x}_i)] \\ &= \alpha \left\{ E[y_i x_i] - [\overline{y}_i \overline{x}_i] \right\} \end{aligned} \tag{3.23}$$

โดย

E คือ Statistical expectation operator

เมื่อพิจารณา (3.19) และ (3.23) จะเห็นว่าพจน์แรกต่างกันที่ (3.23) ใช้ตัวปฏิบัติการ E กับผลคูณระหว่างข้อมูลขาเข้ากับข้อมูลขาออกแทนที่จะเป็นค่าผลคูณนั้นโดยตรง สำหรับพจน์ที่ 2 ของ (3.23) $\bar{y}_i\bar{x}_i$ สามารถพิจารณาเป็นค่าขีดจำกัดที่ป้อนให้กับนิวรอน

2. กฎการเรียนรู้แบบปรับค่าผิดพลาด

การเรียนรู้แบบปรับความผิดพลาดนับว่าเป็นการเรียนรู้ที่สำคัญวิธีหนึ่ง ค่าผลต่างระหว่างข้อมูลขาออกที่ต้องการจะได้ d_i และข้อมูลขาออกที่ได้รับจากโครงข่ายจริงๆ y_i เรียกว่าค่าผิดพลาด e_i นั่นคือ

$$e_i = d_i - y_i \tag{3.24}$$

เป้าหมายของการเรียนรู้แบบปรับค่าผิดพลาด คือ การหาค่าต่ำสุดของฟังก์ชันเป้าหมาย (Cost function) ซึ่งเป็นฟังก์ชันของค่าผิดพลาด เมื่อเลือกฟังก์ชันเป้าหมายที่แทนค่าผิดพลาดโดยรวมอย่างเหมาะสม ก็จะได้ผลลัพธ์ที่ดี สัญญาณการเรียนรู้ l มีค่าดังนี้

$$l \triangleq \left[d_i - f\left(w_i^T x\right)\right] \dot{f}\left(w_i^T x\right) \tag{3.25}$$

โดยที่

$\dot{f}\left(w_i^T x\right)$ คือ อนุพันธ์ของฟังก์ชันกระตุ้นเทียบกับ w_i

จะสังเกตได้ว่าการเรียนรู้แบบปรับค่าผิดพลาดเป็นการเรียนรู้แบบมีการควบคุมและฟังก์ชันกระตุ้น ต้องเป็นฟังก์ชันที่สามารถหาอนุพันธ์ได้ เมื่อใช้ฟังก์ชันเป้าหมาย E ที่เป็นฟังก์ชันของค่าผิดพลาดกำลังสองนั่นคือ

$$E = \tfrac{1}{2}\left[d_i - f\left(w_i^T x\right)\right]^2 = \tfrac{1}{2}\left[d_i - y_i\right]^2 \tag{3.26}$$

เมื่อหาค่าการเกรเดียนท์ (Gredient) ของ E เพื่อหาค่าต่ำสุด นั่นคือ

$$\nabla E = -(d_i - y_i)\dot{f}\left(w_i^T x\right)x \tag{3.27}$$

แต่ละค่าของเวกเตอร์ของค่าเกรเดียนท์สามารถหาได้จาก

$$\frac{\partial E}{\partial w_{ij}} = -(d_i - y_i)\dot{f}\left(w_i^T x\right)x_j \tag{3.28}$$

เนื่องจาก

$$\Delta w_i = -\alpha \nabla E \tag{3.29}$$

นำค่า ∇E จาก (3.27) มาแทน จะได้

$$\Delta w_i = \alpha(d_i - y_i)\dot{f}\left(w_i^T x\right)x \tag{3.30}$$

การเปลี่ยนแปลงของค่าถ่วงน้ำหนักแต่ละค่าหาได้จากความสัมพันธ์

$$\Delta w_{ij} = \alpha(d_i - y_i)\dot{f}\left(w_i^T x\right)x_j \tag{3.31}$$

การเลือกใช้ฟังก์ชันเป้าหมายดัง (3.26) สามารถเรียกได้ว่าการเรียนรู้แบบผลต่างๆ (Delta learning rule) หรือการเรียนรู้แบบเพอร์เซ็ปตรอนสำหรับฟังก์ชันกระตุ้นแบบต่อเนื่อง (Continuous perceptron learning rule)

ในกรณีที่ $f\left(w_i^T x\right) = w_i^T x$ สามารถเรียกได้ว่าการเรียนรู้แบบ Windrow– Hoff หรือการเรียนรู้แบบ LMS (Least mean square learning rule) นั่นคือจะได้สัญญาณการเรียนรู้

$$l \triangleq d_i - w_i^T x \tag{3.32}$$

สามารถหาค่าการเปลี่ยนแปลงของเวคเตอร์ของค่าถ่วงน้ำหนักได้ดังนี้

$$\Delta w_{ij} = \alpha(d_i - w_i^T x)x \tag{3.33}$$

การเปลี่ยนแปลงของค่าถ่วงน้ำหนักแต่ละค่าหาได้จากความสัมพันธ์

$$\Delta w_{ij} = \alpha(d_i - w_i^T x)x_j \quad ; j = 1, 2, ..., n \tag{3.34}$$

การปรับแก้ค่าถ่วงน้ำหนักด้วยวิธี Levenberg-Marquardt Backpropagation Levenberg

$$JJ = J(w) \times J(w) \tag{3.35}$$

$$JE = J(w) \times E \tag{3.36}$$

$$\Delta w = -(JJ + \mu I) \cdot JE^{-1} \tag{3.37}$$

โดย

J คือ Jacobian matrix

E คือ ฟังก์ชันเป้าหมายที่เป็นฟังก์ชันของค่าผิดพลาดกำลังสอง

I คือ Identity matrix

μ คือ ค่าคงที่การเรียนรู้ค่าหนึ่ง

3. การเรียนรู้แบบคอมแพ็ตทิทีฟ

การเรียนรู้แบบคอมแพ็ตทิทีฟนี้จะมีนิวรอนที่เหมาะสมที่สุดในกลุ่มเพียงหนึ่งเดียวเท่านั้นที่จะได้รับการกระตุ้น กล่าวคือ ณ เวลาหนึ่งจะมีเพียงนิวรอนที่ได้รับการคัดเลือกเท่านั้น ที่จะมีการเรียนรู้ เราอาจจะเรียกว่าการเรียนรู้แบบผู้ชนะ (Winner-take –all learning) นิวรอนที่ได้รับการฝึกสอนคือนิวรอนที่มีความเหมาะสมในกลุ่ม หรืออาจเป็นนิวรอนที่เหมาะสมที่สุด นั่นคือก็อาจมีนิวรอนเพียงหนึ่งเดียว หรือหลายนิวรอนที่จะมีการเรียนรู้ในเวลาหนึ่ง เวกเตอร์ค่าถ่วงน้ำหนักของนิวรอนที่เหมาะสม (Winning neuron)

$$w_m = [w_{m1}\ w_{m2}\ \ldots\ w_{mn}]^T \tag{3.38}$$

เวกเตอร์ค่าถ่วงน้ำหนักเปลี่ยนแปลง ดังนี้

$$\Delta w_m = \alpha\left(x - w_m\right) \tag{3.39}$$

การเปลี่ยนแปลงของค่าถ่วงน้ำหนักแต่ละค่า หาได้จากความสัมพันธ์

$$\Delta w_{mj} = \alpha\left(x_j - w_{mj}\right)\ ;\ j = 1, 2, \ldots, n \tag{3.40}$$

สำหรับเวกเตอร์ของค่าถ่วงน้ำหนักนิวรอนอื่นๆจะไม่เปลี่ยนแปลง นั่นคือ

$$\Delta w_k = 0 \ ; k \neq m \tag{3.41}$$

ในการเลือกนิวรอนที่เหมาะสมนั้นจะใช้ระดับกระตุ้นเป็นเกณฑ์การตัดสินใจ ตัวอย่างเช่น

$$w_m^T = \max\left(w_i^T x\right) \ ; i = 1, 2, ..., p \tag{3.42}$$

โดยที่

p คือ จำนวนนิวรอนทั้งหมด

การเรียนรู้แบบคอมแพ็ตทิทีฟจัดว่าเป็นการเรียนรู้ที่ไม่มีการควบคุม ตัวอย่างการประยุกต์ใช้งานได้แก่ใช้ในการจำแนกกลุ่มข้อมูลขาเข้า หรือใช้กับโครงข่ายแบบจัดตัวเอง (Self-organizing network) เป็นต้น

4. กฎการเรียนรู้แบบโบลท์มานน์

การเรียนรู้แบบโบลท์มานน์เป็นการเรียนรู้ที่ต่างจากที่กล่าวมา เนื่องจากเป็นขั้นตอนวิธีการเรียนรู้แบบสโตซาสติก ซึ่งประยุกต์มาจากทฤษฎีทางเทอร์โมไดนามิกส์ โครงข่ายแบบโบลท์มานน์ซึ่งใช้การเรียนรู้แบบโบลท์มานน์ เป็นโครงข่ายแบบย้อนกลับสถานะของนิวรอนจะเป็นเปิด/ปิด (On-off) คุณสมบัติของโครงข่ายถูกกำหนดด้วยฟังก์ชันลังงาน (Energy function) E ซึ่งเป็นฟังก์ชั่นที่บอกความสัมพันธ์ระหว่างนิวรอน ฟังก์ชันพลังงานมีรูปสมการดังนี้

$$E = -\frac{1}{2}\sum_i \sum_{\substack{j \\ j \neq i}} w_{ij} s_i s_j \tag{3.43}$$

โดยที่

s_i คือ สถานะของนิวรอน i

w_{ij} คือ ค่าถ่วงน้ำหนักระหว่างนิวรอน i กับ นิวรอน j

จาก (3.43) จะไม่มีการป้อนกลับมาให้ตัวมันเอง ($i \neq j$) การทำงานเริ่มจากการสุ่มเลือกนิวรอนเพื่อทำการฝึกสอน การเปลี่ยนสถานะของนิวรอน i จาก ปิดไปเป็นเปิด หรือจากเปิดไปเป็นปิด ที่อุณหภูมิ T มีค่าความน่าจะเป็นดังนี้

$$w(s_i \rightarrow -s_i) = \frac{1}{1+\exp(-\Delta E_i / T)} \tag{3.44}$$

โดยที่

ΔE คือ พลังงานที่เปลี่ยนแปลง

นิวรอนของโครงข่ายแบบโบลท์มานน์แบ่งออกเป็นนิวรอนที่มองเห็น (Visible neuron) และนิวรอนซ่อน นิวรอนที่มองเห็นจะเชื่อม (Interface) ระหว่างโครงข่ายประสาทกับภายนอก การทำงานของโครงข่ายมี 2 สภาวะ คือ

1) ภาวะมีเงื่อนไข (Clamped condition) ทำให้นิวรอนที่มองเห็นถูกควบคุมจากภายนอก
2) ภาวะอิสระ (Free running condition) ยอมให้นิวรอนทั้งสองแบบทำงานโดยอิสระนิวรอนซ่อนจะทำงานที่ภาวะอิสระเสมอ

การเปลี่ยนแปลงค่าถ่วงน้ำหนักแต่ละค่าหาได้จากความสัมพันธ์

$$\Delta w_{ij} = \alpha\left(PC_{ij} - PF_{ij}\right) \tag{3.45}$$

โดยที่

PC_{ij} คือ ค่าความน่าจะเป็นที่ s_i และ s_j จะอยู่ในภาวะเปิดเมื่อนิวรอนที่มองเห็นอยู่ในภาวะมีเงื่อนไขขณะที่โครงข่ายอยู่ในสถานะสมดุล

PF_{ij} คือ ความน่าจะเป็นที่ s_i และ s_j จะอยู่ในภาวะเปิด เมื่อทุกนิวรอนอยู่ในภาวะอิสระขณะที่โครงข่ายอยู่ในสถานะสมดุล

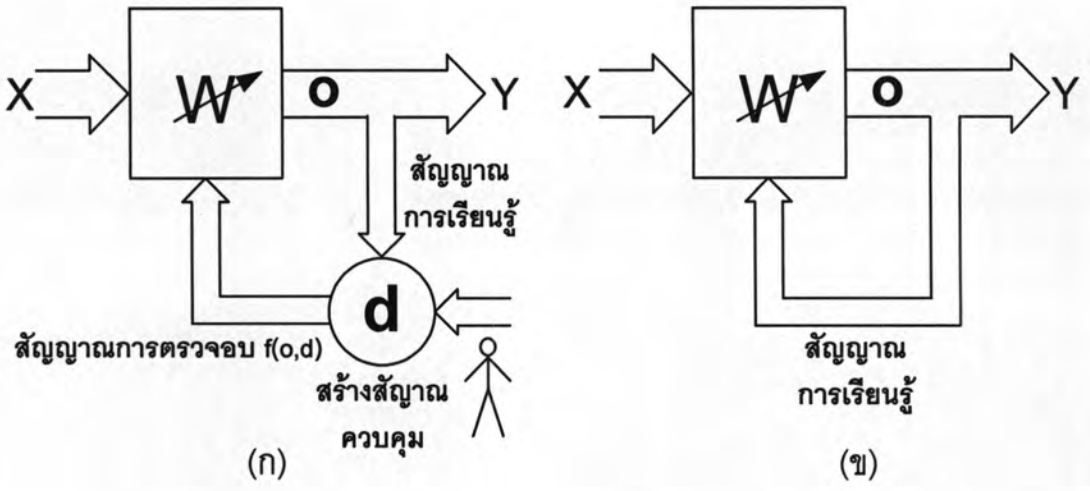
การเรียนรู้แบบโบลท์มานน์ จะผ่านกระบวนการทำซ้ำจนกระทั่งจุดที่เรียกว่า สมดุลทางอุณหภูมิ

3.4.2 รูปแบบการเรียนรู้

รูปแบบการเรียนรู้ออกเป็น 3 รูปแบบคือ

1. การเรียนรู้แบบมีการควบคุม (Supervised learning)
2. การเรียนรู้แบบไม่มีการควบคุม (Unsupervised learning)
3. การเรียนรู้แบบเสริม (Reinforcement learning)

I14717586



รูปที่ 3.11 รูปแบบการเรียนรู้ (ก) แบบมีการควบคุม (ข) แบบไม่มีการควบคุม

1. การเรียนรู้แบบมีการควบคุม

การเรียนรู้แบบนี้จะทราบข้อมูลขาออกที่ต้องการจากรูปที่ 3.11ตัวบอกความสัมพันธ์ระหว่างข้อมูลขาออกของโครงข่ายกับข้อมูลขาออกที่ต้องการ (o, d relationship generator) ทำหน้าที่เป็นครูที่คอยควบคุมการเรียนรู้ ข้อมูลขาออกที่ต้องการคือ ข้อมูลขาออกที่ต้องการคือข้อมูลที่ครูใช้เปรียบเทียบกับผลตอบสนองที่ได้รับ ผลที่ได้ (f(o, d)) จะใช้ในการปรับทิศทางการเรียนรู้ให้เหมาะสม จากพฤติกรรมดังกล่าวทำให้โครงข่ายประสาทที่มีการเรียนรู้แบบมีการควบคุมจึงต้องมีชุดข้อมูลขาเข้าและส่งออก เราเรียกว่า ชุดการฝึกสอน (Training set หรือ Training data)

2. การเรียนรู้แบบไม่มีการควบคุม

การเรียนรู้แบบไม่มีการควบคุมต่างจากแบบมีการควบคุม เนื่องจากเราไม่ทราบข้อมูลขาออกที่ต้องการ การเรียนรู้จึงใช้ข้อมูลขาเข้า ภาวะแวดล้อม และตัวโครงข่ายเองในการปรับเข้าหาคำตอบ หากจะเปรียบเทียบกับการเรียนรู้จริงก็คือการศึกษาด้วยตนเองนั่นเอง

3. การเรียนรู้แบบเสริม

การเรียนรู้แบบเสริมเหมือนกับการเรียนรู้แบบมีการควบคุม [4] ยกเว้นข้อมูลขาออกแทนที่จะได้ข้อมูลขาออกที่ถูกต้องสำหรับข้อมูลขาเข้าแต่ละชุด การเรียนรู้แบบเสริมนี้จะให้ค่าข้อมูลขาออกเป็นระดับหรือคะแนนที่บอกพฤติกรรมของโครงข่ายที่มีต่อข้อมูลขาเข้าแต่ละชุด การเรียนรู้แบบเสริมสามารถนำไปประยุกต์ใช้กับงานระบบควบคุมได้เป็นอย่างดี

3.4.3 การนอร์มอลไลซ์ (Normalization)

ชุดข้อมูลขาเข้าของโครงข่ายที่ใช้ในการประมาณค่า ซึ่งตัวแปรในข้อมูลขาเข้าบางครั้งนั้นมีช่วงข้อมูลที่ต่างกันมาก ค่าเหล่านี้จึงจำเป็นต้องทำให้อยู่ในช่วงเดียวกัน หรือที่เรียกว่าการนอร์มอลไลซ์ โดย (3.46)

$$y = (y_{\max} - y_{\min})\frac{(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \tag{3.46}$$

เมื่อ

y คือ ค่าข้อมูลฝึกระบบที่นอร์มอลไลซ์แล้ว

$y_{\min}$ คือ ค่าข้อมูลต่ำสุดที่

$y_{\max}$ คือ ค่าข้อมูลสูงสุดที่ต้องการนอร์มอลไลซ์

x คือ ค่าข้อมูลฝึกระบบ

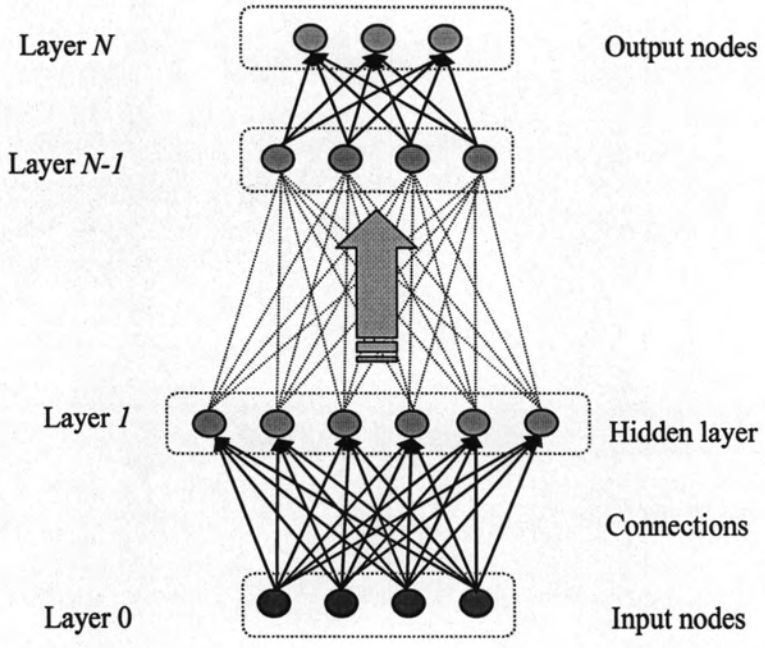
$x_{\min}$ คือ ค่าต่ำสุดของชุดข้อมูลฝึกระบบ

$x_{\max}$ คือ ค่าสูงสุดของชุดข้อมูลฝึกระบบ

โดยปกติแล้ว $y_{\max}$ = 1 และ $y_{\min}$ = -1 เพื่อรองรับกระบวนการของโครงข่ายประสาทเทียม

3.5 โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น

โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron Neural Network: MLPNN) เป็นโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้า (Feed-forward neural network) ประกอบด้วย ปมประสาท (Neuron) ที่เรียกว่า เพอร์เซ็ปตรอน (Perceptron) โดยที่เพอร์เซ็ปตรอนประกอบด้วยปมประสาทเพียงหนึ่งยูนิตที่มีตัวถ่วงน้ำหนักการเชื่อมโยง (Weight) และค่าไบแอส (Bias) โครงข่ายประสาทเทียมเพอร์เซ็ปตรอนหลายชั้นสามารถสร้างได้จากการนำเพอร์เซ็ปตรอนมาต่อกันเป็นโครงข่าย ดังในรูปที่ 3.12 การเรียนรู้ ของโครงข่ายก็คือ การหาค่าถ่วงน้ำหนักที่เหมาะสมโครงข่ายนั้นๆนั้นๆ [4] การการนำโครงข่ายประสาทเทียมมา ใช้งานต้องมี 2 ขั้นตอน คือ ขั้นตอนการเรียนรู้ (Learning) และขั้นตอนการทดสอบ (Testing) ซึ่งในการเรียนรู้นั้นมี 2 ลักษณะ คือ การคำนวณไปข้างหน้า (Forward propagation) และการคำนวณแพร่กลับ (Backward propagation) ส่วนแรกการคำนวณไปข้างหน้าเป็นการป้อนข้อมูลขาเข้าให้กับโครงข่าย จากนั้นโครงข่ายจะทำการคำนวณจากชั้นขาเข้าไปยังชั้นขาออก



รูปที่ 3.12 โครงข่ายประสาทเทียมแบบหลายชั้น ที่มีกระบวนการเรียนรู้แบบแพร่กลับ