

บทที่ 3

ทฤษฎีโครงข่ายประสาทเทียม

3.1 บทนำ

โครงข่ายประสาทเทียม (Artificial Neural Network, ANN) เป็นแขนงหนึ่งของปัญญาประดิษฐ์ (Artificial Intelligence, AI) ซึ่งมีโครงสร้างเชิงสถาปัตยกรรม และการทำงานคล้ายคลึงกับการทำงานของเซลล์สมอง หรือระบบประสาทของมนุษย์ โดยสามารถเรียนรู้และปรับตัวตามข้อมูลที่ได้รับจากสภาพแวดล้อมภายนอก และให้คำตอบในปัญหาที่ซับซ้อนได้น่าเชื่อถือ [35]

เนื่องจากคุณสมบัติที่สำคัญคือความยืดหยุ่นในการจำลองความสัมพันธ์ระหว่างปัจจัยนำเข้า (Input) กับผลลัพธ์ (Output) ที่มีความซับซ้อน และสามารถตอบสนองกับข้อมูลที่ไม่เคยเห็นได้ดี ทำให้มีการนำทฤษฎีโครงข่ายประสาทเทียมมาประยุกต์ใช้แก้ปัญหาทางจริงหลายด้านอย่างกว้างขวาง เช่น การเงินการธนาคาร, อวกาศ, ระบบป้องกันประเทศ, ระบบรักษาความปลอดภัย, การแพทย์, ระบบสื่อสาร, ระบบขนส่ง, การบันเทิง รวมทั้งงานทางด้านวิศวกรรม [36, 37] โดยนำโครงข่ายประสาทเทียมมาใช้ในการจดจำรูปแบบ (Recognition), การจัดกลุ่ม (Clustering), การจัดหมวดหมู่ (Classification) และการพยากรณ์ (Forecasting) เป็นต้น ในการศึกษาวิจัยนี้ได้นำโครงข่ายประสาทเทียมแบบป้อนไปข้างหน้าโดยใช้การเรียนรู้แบบแพร่ค่าย้อนกลับ (Feedforward Network with Back Propagation Algorithm) มาใช้ในการพยากรณ์กำลังรับแรงอัดซึ่งจะได้กล่าวต่อไป

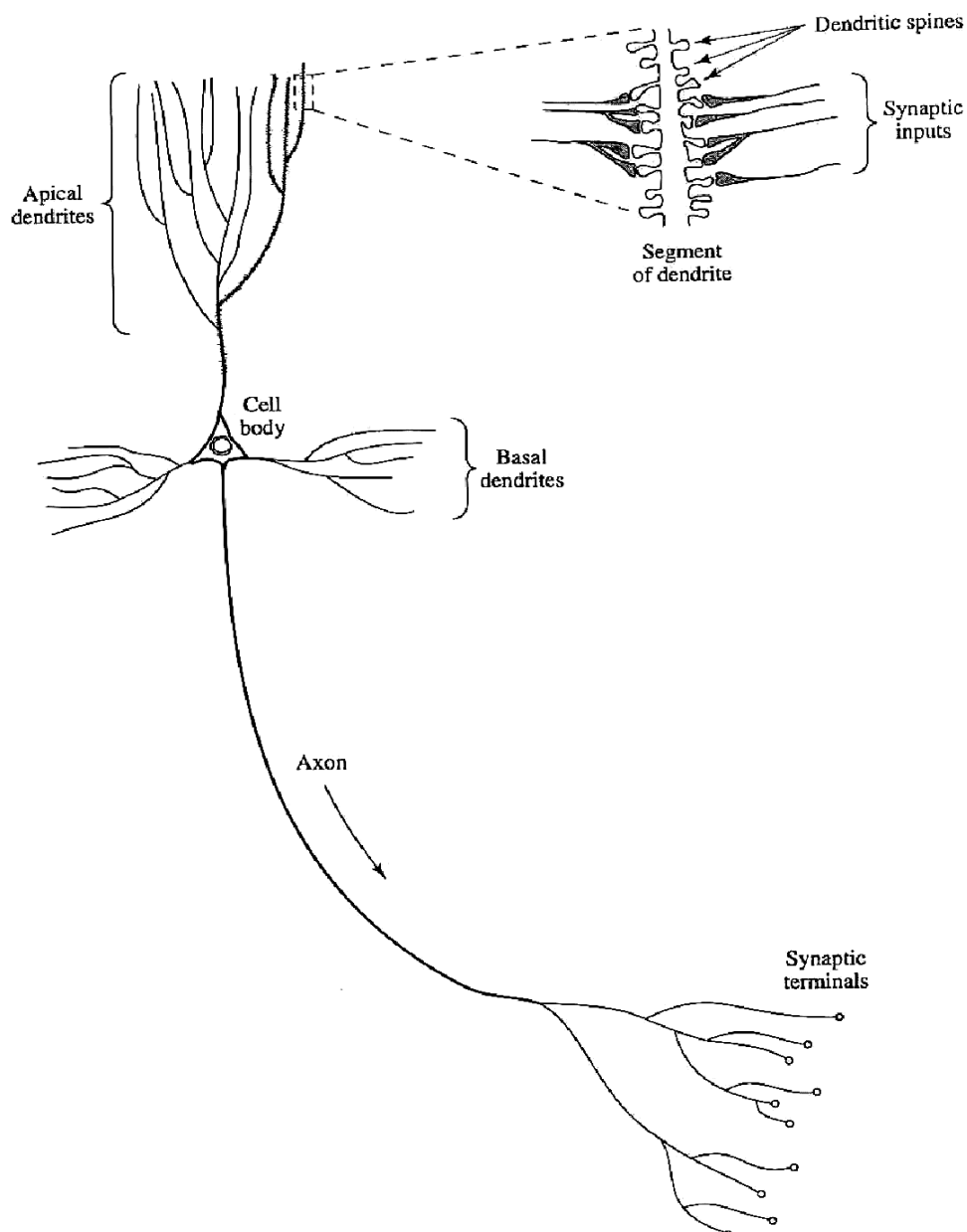
3.2 การจำลองเซลล์ประสาท

3.2.1 เซลล์ประสาทชีวภาพ

ระบบคิดคำนึงของมนุษย์มีโครงสร้างพื้นฐานจากเซลล์สมองที่เรียกว่าเซลล์ประสาท (Neuron) เรียงเป็นชั้นๆ จำนวนมากประมาณ 10^{11} ยูนิต และมีการเชื่อมโยงภายในระหว่างเซลล์ประมาณ 10^5 ยูนิต ในแต่ละเซลล์ประสาทจะมีคุณลักษณะแตกต่างกันออกไปโดยมีการทำงานคล้ายกันคือ รับสัญญาณเข้า, ประมวลผล และส่งสัญญาณไฟฟ้าเคมีผ่านไปยังเซลล์ประสาทอื่น

จากรูปที่ 3.1 เซลล์ประสาทประกอบด้วยเดนไดรต์ (Dendrites) ซึ่งเป็นส่วนแขนงที่ขยายออกจากตัวเซลล์เพื่อรับสัญญาณจากเซลล์ประสาทอื่นที่ส่งมาทางจุดเชื่อมต่อที่เรียกว่าไซแนปส์ (Synapse) และปลายในการส่งกระแสประสาทเรียกว่าแอกซอน (Axon) เซลล์ประสาททำงานด้วยปฏิกิริยาไฟฟ้าเคมี เมื่อมีการกระตุ้นด้วยสิ่งเร้าภายนอกหรือกระตุ้นด้วยเซลล์ด้วยกัน กระแส

ประสาทจะวิ่งผ่านเดนไดรต์เข้าสู่ตัวเซลล์ (Cell Body) ซึ่งจะเป็นตัวตัดสินใจว่าต้องกระตุ้นเซลล์อื่นๆ ต่อหรือไม่ ถ้ากระแสประสาทแรงพอตัวเซลล์ก็จะกระตุ้นเซลล์อื่นโดยส่งผ่านทางแอกซอน

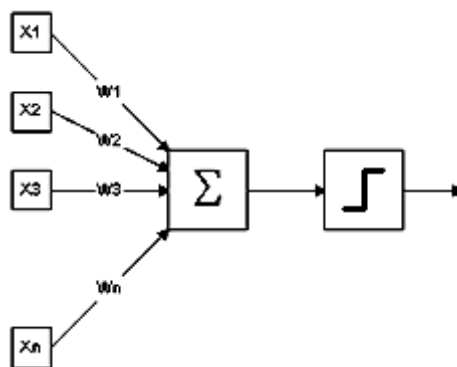


รูปที่ 3.1 โครงสร้างเซลล์ประสาทชีวภาพ [35]

3.2.2 เซลล์ประสาทเทียม

แบบจำลองของเซลล์ประสาททางชีวภาพ หรือที่เรียกว่าเซลล์ประสาทเทียม (Artificial Neuron) แสดงได้ดังรูปที่ 3.2 โดยประกอบด้วยโครงสร้างพื้นฐานที่สำคัญคือ หน่วย (Node) หรือ ยูนิต (Unit), ตัวแปรด้านเข้า (Input), ตัวแปรด้านออก (Output) และค่าถ่วงน้ำหนัก (Weight) ซึ่งสามารถสรุปความสัมพันธ์ระหว่างเซลล์ประสาททางชีววิทยาและเซลล์ประสาทเทียมดังนี้

- การประมวลผล จะเกิดขึ้นในหน่วยประมวลผลย่อยคือ หน่วย (Node) หรือ ยูนิต (Unit) ซึ่งจำลองมาจากลักษณะการทำงานของตัวเซลล์
- การส่งสัญญาณระหว่างหน่วยด้วยส่วนที่เชื่อมติดกัน จำลองมาจากการเชื่อมต่อของเดนไดรต์ และแอกซอน
- แต่ละการเชื่อมต่อประกอบด้วยค่าน้ำหนักที่แตกต่างกัน โดยขึ้นอยู่กับอิทธิพลที่หน่วยจะได้รับจากหน่วยอื่นๆ ซึ่งจำลองมาจากไซแนปส์ โดยค่าน้ำหนักที่ได้รับจะทำหน้าที่เสมือนความรู้ที่ถูกรวบรวมไว้ใช้แก้ปัญหาเฉพาะอย่างของมนุษย์
- ภายในหน่วยมีฟังก์ชันที่ใช้ในการกำหนดสัญญาณด้านออกที่เรียกว่า ฟังก์ชันถ่ายโอน (Transfer Function) หรือ ฟังก์ชันกระตุ้น (Activation Function)



รูปที่ 3.2 โครงสร้างเซลล์ประสาทเทียม

3.3 โครงข่ายประสาทเทียมแบบไปข้างหน้า

การจัดประเภทโครงสร้างโครงข่ายประสาทเทียมตามทิศทางการเคลื่อนที่ของสัญญาณแบ่งได้ 2 ชนิดคือ โครงข่ายแบบไปข้างหน้า (Feedforward Network) และ โครงข่ายแบบส่งค่ากลับ (Recurrent Network) ซึ่งในบทนี้จะกล่าวถึงเฉพาะโครงข่ายประสาทเทียมแบบไปข้างหน้า

โครงข่ายประสาทเทียมแบบไปข้างหน้าจะประกอบด้วยชั้นดังต่อไปนี้

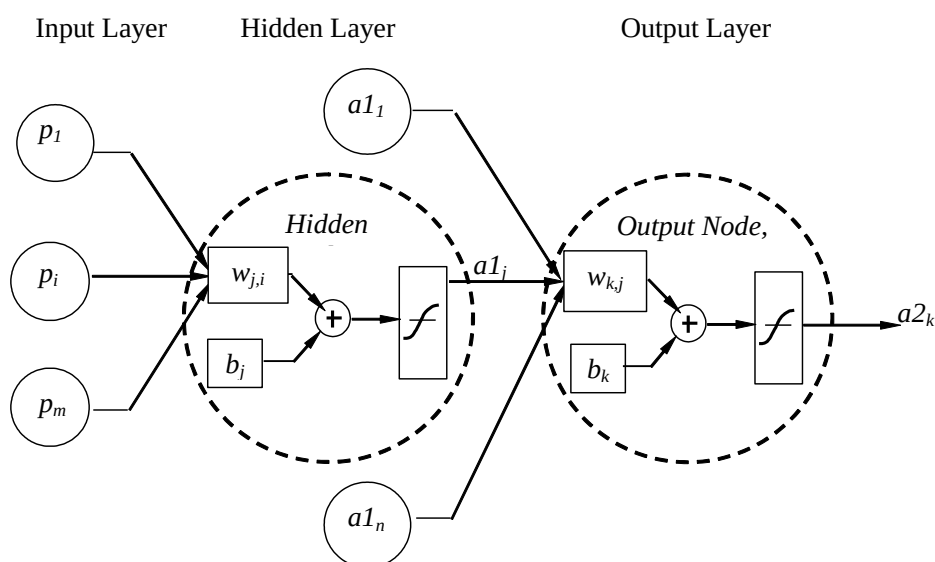
1. ชั้นรับข้อมูล (Input Layer)
2. ชั้นซ่อน (Hidden Layer) ซึ่งในโครงข่ายจะมีหรือไม่มีชั้นซ่อนก็ได้ ถ้าไม่มีชั้นซ่อนเรียก

โครงข่ายว่า Single-Layer Feedforward Network ส่วนโครงข่ายที่มีชั้นซ่อนเรียกว่า Multilayer Feedforward Network ซึ่งจำนวนชั้นซ่อนอาจมีจำนวน 1 ชั้น หรือมากกว่า

3. ชั้นแสดงผล (Output Layer)

โดยในแต่ละชั้นจะประกอบด้วย หน่วย (Node) ซึ่งจะทำหน้าที่ประมวลผล และเชื่อมต่อผลลัพธ์ไปยังชั้นถัดไป การเชื่อมต่อของแต่ละชั้นจะเป็นแบบถึงกันหมดจึงทำให้สัญญาณข้อมูล

เคลื่อนที่ไปในทิศทางเดียวคือเคลื่อนที่ไปข้างหน้า จากชั้นรับข้อมูลไปยังชั้นแสดงผล โดยจำนวนชั้นแฝง และหน่วยในแต่ละชั้นแฝงจะขึ้นอยู่กับความซับซ้อนของปัญหา ส่วนจำนวนหน่วยในชั้นรับข้อมูลจะขึ้นกับจำนวนตัวแปรที่ส่งผลต่อค่าที่จะพยากรณ์ และจำนวนหน่วยในชั้นแสดงผลคือผลลัพธ์ที่ต้องการจากโครงข่าย [38] โครงสร้างเบื้องต้นของโครงข่ายประสาทเทียม แบบ Feed-Forward แสดงในรูปที่ 3.3



รูปที่ 3.3 โครงข่ายประสาทเทียมแบบไปข้างหน้า (Feedforward Network)

ค่าผลลัพธ์จากแต่ละหน่วย จะสามารถหาได้จากสมการที่ 3.1 และ 3.2 ดังนี้

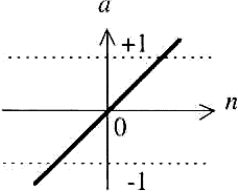
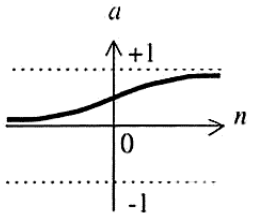
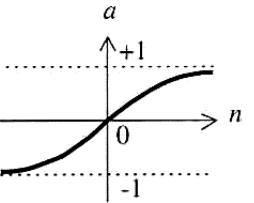
$$a1_j = f\left(\sum_{i=1}^m (w_{j,i} \cdot p_i) + b_j\right) \quad (3.1)$$

$$a2_k = f\left(\sum_{j=1}^n (w_{k,j} \cdot a1_j) + b_k\right) \quad (3.2)$$

- โดยที่ p_i คือ ปัจจัยนำเข้า (Input)
 $a1_j$ คือ ค่าผลลัพธ์ที่ได้จากการประมวลผลของ ชั้นซ่อน (Hidden Layer)
 $a2_k$ คือ ค่าผลลัพธ์จากชั้นแสดงผล
 w คือ ค่าน้ำหนัก (Weight)
 b คือ ค่าเบี่ยงเบน (Bias)
 $f(c)$ คือ ฟังก์ชันถ่ายโอน (Transfer Function) หรือ ฟังก์ชันกระตุ้น (Activation Function)

ฟังก์ชันถ่ายโอน (Transfer Function) หรือ ฟังก์ชันกระตุ้น (Activation Function) จะเป็นตัวกำหนดค่าผลลัพธ์ของโครงข่ายประสาทเทียม ซึ่งฟังก์ชันถ่ายโอนมีอยู่หลายชนิด การที่จะเลือกฟังก์ชันชนิดใดขึ้นอยู่กับรูปแบบปัญหา, ลักษณะของผลลัพธ์ว่าเป็นค่าต่อเนื่องหรือไม่, ขอบเขตของผลลัพธ์ เป็นต้น แต่ในที่นี้จะเสนอเฉพาะฟังก์ชันถ่ายโอนที่ใช้ในการศึกษานี้เท่านั้นสรุปได้ดังตารางที่ 3.1

ตารางที่ 3.1 ชนิดของฟังก์ชันถ่ายโอน (Transfer Function)

ชนิดของฟังก์ชันถ่ายโอน	รูปแบบ	สมการฟังก์ชันถ่ายโอน	ขอบเขตของ $a(n)$
Linear Transfer Function		$a(n) = n$	$[-\infty, \infty]$
Log-Sigmoid Transfer Function		$a(n) = \frac{1}{1 + e^{(-n)}}$	$[0, 1]$
Tan-Sigmoid Transfer Function		$a(n) = \frac{1 - e^{(-n)}}{1 + e^{(-n)}}$	$[-1, 1]$

ส่วนค่าน้ำหนัก (Weight) และค่าเบี่ยงเบน (Bias) เป็นค่าที่ได้จากกระบวนการเรียนรู้ โดยกระบวนการเรียนรู้ที่นิยมใช้กับโครงข่ายประสาทเทียมแบบไปข้างหน้าคือการเรียนรู้แบบแพร่ค่าย้อนกลับ (Backpropagation Algorithm) [38]

3.4 การเรียนรู้แบบแพร่ค่าย้อนกลับ

คุณสมบัติสำคัญของโครงข่ายประสาทเทียม คือ ความสามารถในการเรียนรู้จากตัวอย่าง โดยพยายามคำนวณหาความสัมพันธ์ระหว่างปัจจัยนำเข้า (Input) และผลลัพธ์ (Output) การเรียนรู้นี้จะเริ่มจากการสุ่มค่าน้ำหนัก (Weight) และค่าเบี่ยงเบนเริ่มต้น (Bias) ค่าผลลัพธ์ที่ได้จากค่าเริ่มต้นจะถูกนำมาเปรียบเทียบกับผลลัพธ์จริง ค่าที่แตกต่างกันจะถูกนำมาปรับค่าน้ำหนักและค่าเบี่ยงเบนโดย

วิธีลองผิดลองถูก (Trial and Error) จนได้ผลลัพธ์ที่ใกล้เคียงหรือตรงกับผลลัพธ์จริง ค่าน้ำหนักและค่าเบี่ยงเบนสุดท้ายจะถูกนำมาใช้ในการพยากรณ์ผลลัพธ์ที่เกิดขึ้นจากข้อมูล (Input) ใหม่

กระบวนการในการปรับค่าน้ำหนักและค่าเบี่ยงเบนที่นิยมใช้ คือ การเรียนรู้แบบแพร่ค่าย้อนกลับ (Backpropagation) ซึ่งจะประกอบด้วยสัญญาณ 2 ส่วน ดังนี้

1. สัญญาณส่ง (Function Signals) คือ ข้อมูลหรือผลลัพธ์ของแต่ละหน่วยที่ถูกส่งเข้าสู่โครงข่ายประสาทเทียม

2. สัญญาณแก้ไข (Error Signals) คือ ค่าที่แตกต่างระหว่างผลลัพธ์ที่ได้จากการคำนวณและผลลัพธ์จริงที่ถูกส่งกลับสู่โครงข่ายประสาทเทียม เพื่อปรับค่าน้ำหนัก และค่าเบี่ยงเบน

ขั้นตอนการปรับค่าน้ำหนักและค่าเบี่ยงเบนมีอยู่หลายวิธีด้วยกัน เช่น Gradient Descent, Gradient Descent with Adaptive Learning Rate, Gradient Descent with Momentum, Gradient Descent with Momentum and Adaptive Learning Rate, Resilient, Bayesian Regularization และ Levenberg-Marquardt เป็นต้น โดยในการศึกษานี้จะใช้วิธี Levenberg-Marquardt Algorithm ในการปรับค่าน้ำหนัก และค่าเบี่ยงเบน เนื่องจากวิธีนี้เหมาะสมในการแก้ไขปัญหาประเภทการประมาณค่า นอกจากนั้นยังเป็นวิธีที่มีความรวดเร็วที่สุดในการปรับค่า [39] โดยมีขั้นตอนดังนี้

$$x_{(new)} = x_{(old)} + \Delta x$$

$$x_{(new)} = x_{(old)} - [J^T \cdot J + \mu I]^{-1} \cdot J^T \cdot e \quad (3.3)$$

$$g = 2J^T \cdot e \quad (3.4)$$

โดยที่ x คือ ค่าถ่วงน้ำหนัก (Weight) และค่าเบี่ยงเบน (Bias)

J คือ จาคอบีเยนเมตริกซ์ (Jacobian Matrix) ของอนุพันธ์ของค่าผิดพลาด (แต่ละค่า) ต่อค่า x (แต่ละค่า)

J^T คือ Transpose Matrix ของ J

g คือ Gradient Factor

e คือ เมตริกซ์ของค่าผิดพลาด (Error) ระหว่างคำตอบ (Target) กับค่าที่โครงข่ายประสาทเทียมคำนวณได้ (Output)

I คือ Identity Matrix

μ คือ Mu Factor หรือ Learning Factor

จาโคเบียนเมตริกซ์ของค่าถ่วงน้ำหนัก (Weight, w)

$$J(w) = \begin{bmatrix} \frac{\partial e_1}{\partial w_{1,1}} & \frac{\partial e_1}{\partial w_{1,2}} & \cdots & \frac{\partial e_1}{\partial w_{1,s2}} \\ \frac{\partial e_2}{\partial w_{2,1}} & \frac{\partial e_2}{\partial w_{2,2}} & \cdots & \frac{\partial e_2}{\partial w_{2,s2}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial e_{s3}}{\partial w_{s3,1}} & \frac{\partial e_{s3}}{\partial w_{s3,2}} & \cdots & \frac{\partial e_{s3}}{\partial w_{s3,s2}} \end{bmatrix} \quad (3.5)$$

จาโคเบียนเมตริกซ์ของค่าเบี่ยงเบน (Bias, b)

$$J(b) = \begin{bmatrix} \frac{\partial e_1}{\partial b_1} \\ \frac{\partial e_2}{\partial b_2} \\ \vdots \\ \frac{\partial e_{s3}}{\partial b_{s3}} \end{bmatrix} \quad (3.6)$$

เมตริกซ์ของค่าผิดพลาด (Error, e)

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{s3} \end{bmatrix} = \begin{bmatrix} t_1^3 - a_1^3 \\ t_2^3 - a_2^3 \\ \vdots \\ t_3^3 - a_3^3 \end{bmatrix} \quad (3.7)$$

การลู่เข้าของการปรับค่าน้ำหนัก และค่าเบี่ยงเบนจะเร็วหรือช้า นอกจากรูปแบบวิธีการเรียนรู้แล้ว การกำหนดค่าเริ่มต้นของค่าน้ำหนัก และค่าเบี่ยงเบนก็ส่งผลด้วยเช่นกัน

3.5 อภิปราย

โครงข่ายประสาทเทียมเป็นวิธีที่เลียนแบบการทำงานของสมองมนุษย์ซึ่งเหมาะในการแก้ไขปัญหาที่ซับซ้อนวิธีหนึ่ง ด้วยการเรียนรู้ความสัมพันธ์ของปัจจัยนำเข้า (Input) และผลลัพธ์ (Output) จากข้อมูลในอดีต เพื่อค้นหาคำตอบที่เหมาะสมสำหรับปัจจัยนำเข้าชุดใหม่ โดยมีปัจจัยที่ส่งผลต่อประสิทธิภาพของโครงข่ายประสาทเทียมซึ่งประกอบด้วย [35, 38, 40]

1. การเลือกปัจจัยนำเข้า (Input)

การเลือกปัจจัยนำเข้าที่มีความสัมพันธ์กับผลลัพธ์ จะทำให้โครงข่ายประสาทเทียมมีประสิทธิภาพทั้งในกระบวนการเรียนรู้ และการทดสอบ การตรวจสอบความสัมพันธ์ระหว่างปัจจัยนำเข้ากับผลลัพธ์สามารถทำได้โดยการวิเคราะห์สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient)

2. จำนวนชั้นซ่อน และจำนวนหน่วยแต่ละชั้นซ่อน

จำนวนชั้นซ่อน และจำนวนหน่วยแต่ละชั้นซ่อนเป็นปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพของโครงข่ายประสาทเทียม จำนวนชั้นซ่อนและจำนวนหน่วยต้องมีความเหมาะสมไม่มากหรือน้อยเกินไป ถ้ามากเกินไปก็จะเกิดปัญหา Overfitting แต่ถ้าน้อยเกินไปก็จะเกิด Underfitting ซึ่งทั้ง 2 รูปแบบนี้ทำให้ประสิทธิภาพในการพยากรณ์ของโครงข่ายน้อยลง

3. จำนวนข้อมูลที่ใช้ฝึกสอน

หากโครงข่ายประสาทเทียมได้รับข้อมูลในกระบวนการเรียนรู้ที่มาก โครงข่ายจะวิเคราะห์ความสัมพันธ์ภายในขอบเขตเรียนรู้ละเอียดขึ้น ซึ่งส่งผลให้ประสิทธิภาพของโครงข่ายสูงขึ้นตาม

4. การกระจายของข้อมูลที่ใช้ฝึกสอน

โครงข่ายประสาทเทียมจะมีประสิทธิภาพในการหาผลลัพธ์ภายในขอบเขตที่ได้รับการฝึกสอน ถ้าข้อมูลที่ใช้ฝึกสอนมีการกระจายตัวดีภายในขอบเขตที่ศึกษา ประสิทธิภาพในการหาผลลัพธ์ก็จะดีขึ้นด้วย

5. การ Normalization ข้อมูลที่ใช้ฝึกสอน

การจัดระเบียบข้อมูลก่อนที่จะเข้าสู่กระบวนการฝึกสอนโครงข่ายประสาทเทียม จะทำให้กระบวนการฝึกสอนเร็วขึ้นกว่าการนำข้อมูลดิบมาทำการฝึกสอน

6. วิธีการปรับค่าน้ำหนักและค่าเบี่ยงเบน

การปรับค่าน้ำหนักและค่าเบี่ยงเบนมีหลายวิธี ซึ่งแต่ละวิธีก็มีความเหมาะสมกับปัญหาต่างรูปแบบกัน

7. ค่าเริ่มต้นของน้ำหนัก และค่าเบี่ยงเบน

จะมีผลต่อการเข้าใกล้จุด Local Minimum หรือ Global Minimum ของผลลัพธ์ที่ใช้ในกระบวนการฝึกสอน ดังนั้นในการฝึกสอนโครงข่ายประสาทเทียมจึงควรสุ่มเลือกค่าเริ่มต้นและค่าเบี่ยงเบนในแต่ละโครงสร้างหลายๆครั้ง

8. จำนวนรอบการฝึกสอน

จำนวนรอบการฝึกสอนที่เหมาะสมไม่มากหรือน้อยเกินไป จะทำให้ประสิทธิภาพในการหาผลลัพธ์ของโครงข่ายดีขึ้น

9. ชนิดของฟังก์ชันถ่ายโอน

ชนิด รวมทั้งความชันของฟังก์ชันถ่ายโอนจะมีผลต่อกระบวนการเรียนรู้ และผลลัพธ์ที่ได้จากโครงข่ายประสาทเทียม