บทที่ 2

หลักการและทฤษฎีที่เกี่ยวข้อง

ในการจำแนกเสียงร้องของเด็กทารกนั้นได้แบ่งอัลกอริทึมที่ใช้ในการทำงานออกเป็นสี่ส่วนหลักๆคือ สวนแรกคือการสกัดค่าคุณลักษณะ (Feature Extraction) เป็นอัลกอริทึมในการสร้างชุดข้อมูลที่เป็นตัวเลขจากตัวอย่างไฟล์วิดีโอและคลิปเสียง ส่วนที่สองคือ การทำเหมืองข้อมูล (Data Mining) เป็นอัลกอริทึมในการรู้จำเสียงและจำแนกเสียง ส่วนที่สามคือ ภาษาทารกของดันส์ตัน (Dunstan Baby Language) เป็นประเภทของเสียงร้องต่างๆของเด็กทารกที่ได้กำหนดขึ้น และส่วนสุดท้ายคือ การตรวจสอบความสมเหตุสมผลแบบไขว้ (Cross Validation) เป็นอัลกอริทึมทางสถิติในการตรวจสอบประสิทธิภาพของผลลัพธ์ที่ได้

2.1 การสกัดค่าคุณลักษณะ (Feature Extraction)

เป็นขั้นตอนการแปลงเสียงที่เป็นสัญญาณอนาล็อก (Analog Signal) ให้มาเป็นสัญญาณดิจิตอล (Digital Signal) ในรูปของข้อมูลตัวเลข สามารถทำได้โดยอาศัยอัลกอริทึมต่างๆ เช่น สัมประสิทธิ์เซปสตรัมบนสเกลเมล (MFCC) การประมาณพันธะแบบอิงการรับฟังของมนุษย์ (PLP) ราสตา (RASTA) ซึ่งอัลกอริทึมเหล่านี้เป็นอัลกอริทึมที่ได้รับความนิยมในงานด้านการรู้จำเสียง

2.1.1 สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel Frequency Cepatral Coefficients: MFCC)[12]

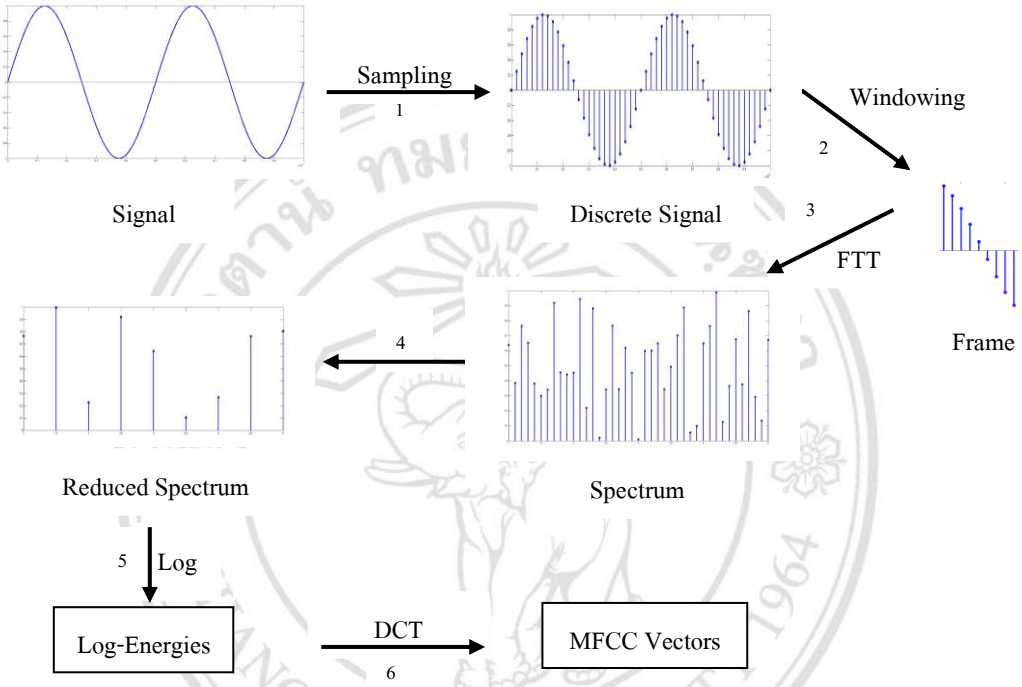
สัมประสิทธิ์เซปสตรัมบนสเกลเมลเป็นอัลกอริทึมที่ได้รับความนิยมอย่างกว้างขวางในการสกัดค่าคุณลักษณะ (Feature) ออกมาจากเสียงของมนุษย์ เนื่องจากโดยปกติแล้วเสียงของมนุษย์นั้นจะอยู่ในช่วงความถี่ที่ต่ำๆ มากกว่าช่วงความถี่ที่สูงๆ ซึ่งอัลกอริทึมนี้จะให้ความสำคัญกับช่วงของความถี่ที่ต่ำโดยทำการขยายช่วงให้กว้างขึ้นและลดช่วงของความถี่ที่สูงลง โดยการปรับความถี่ใหม่นี้เรียกว่า สเกลเมล (Mel Scale) ซึ่งการแปลงความถี่ปกติของเสียงพูดเป็นความถี่ตามสเกลเมลตามสมการ (2.1)

$$f_{mel} = 2595 \log_{10}(1 + \frac{f}{700}) \tag{2.1}$$

เมื่อ f คือความถี่ปกติมีหน่วยเป็นเฮิรตซ์ (Hz)

f_{mel} คือความที่บนสเกลเมล

โดยขั้นตอนการแยกค่าคุณลักษณะออกจากเสียงโดยใช้วิธีสัมประสิทธิ์เซปสตรัมบนสเกลเมล นั้นมีขั้นตอนตามรูปที่ 2.1



รูปที่ 2.1 ขั้นตอนการแปลงเสียงโดยใช้สัมประสิทธิ์เซปสตรัมบนสเกลเมล

ขั้นตอนการทำงานแบ่งออกเป็น 6 ขั้นตอนคือ

ขั้นตอนที่ 1 เริ่มจากนำสัญญาณเสียง (Signal Waveform) มาทำการสุ่มตัวอย่าง (Sampling) เพื่อเปลี่ยนสัญญาณเสียงที่เป็นอนาล็อก (Analog) เป็นสัญญาณดิจิตอล (Digital)

ขั้นตอนที่ 2 นำสัญญาณดิจิตอลทำการแบ่งหน้าต่างแฮมมิ่ง (Hamming Window)

ขั้นตอนที่ 3 นำแต่ละหน้าต่างแฮมมิ่งผ่านกระบวนการแปลงฟูริเยร์ (Fourier Transformation) เพื่อเปลี่ยนสัญญาณเสียงให้อยู่ในโดเมนความถี่ โดยใช้การแปลงฟูริเยร์แบบไม่ต่อเนื่อง (Discrete Fourier Transformation : DFT) หรือการแปลงฟูริเยร์แบบเร็ว (Fast Fourier Transformation : FFT)

ขั้นตอนที่ 4 ทำการปรับความถี่ที่ได้ให้อยู่ในสเกลเมล จากนั้นส่งค่าที่ได้ไปผ่านชุดตัวกรอง (Filter Bank) โดยมีตัวอย่างเป็นตัวกรองสามเหลี่ยม (Triangular Filter) มีสมการตามสมการ (2.2)

$$H_b(b) = \begin{cases} \dfrac{f - CF_{b-1}}{CF_b - CF_{b-1}}; CF_{b-1} < f < CF_b \\ \dfrac{f - CF_{b+1}}{CF_b - CF_{b+1}}; CF_b \le f < CF_{b+1} \\ 0; otherwise \end{cases} \tag{2.2}$$

เมื่อ CF คือความถี่กลางในแบนด์ (Center Frequency of Band)

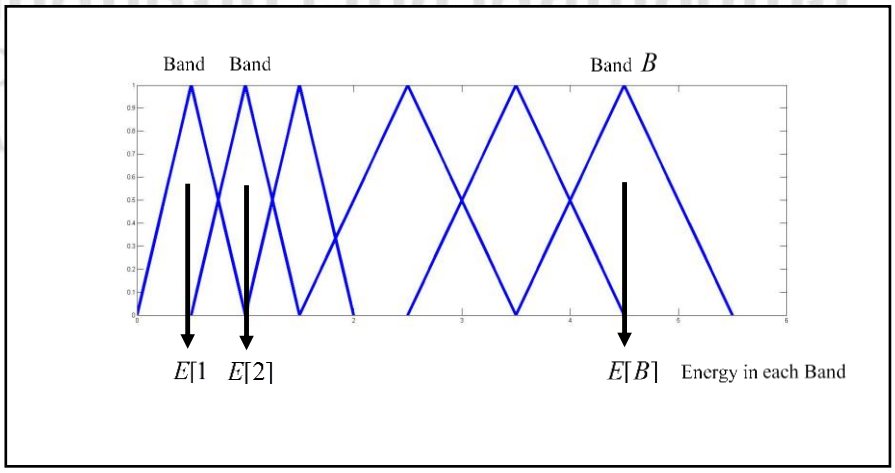
M คือจำนวนแบนด์

b คือลำดับแบนด์

ความถี่กลางในแบนด์ มีหน่วยเป็นเมล (Mel) สามารถหาได้จากสมการ (2.3)

$$CF_b = F_{\min} + b\left(\frac{F_{\max} - F_{\min}}{M + 1}\right) \tag{2.3}$$

ค่าที่ได้จะเป็นสเกลเมลแล้วทำการแปลงเป็นหน่วยเฮิรตซ์ (Hz) ก่อนนำไปผ่านตัวกรอง โดยตัวกรองนั้นมีไว้เพื่อหาค่าพลังงานในแต่ละแบนด์ (Bands) ซึ่งจะหาพลังงานทุกแบนด์ตามรูปที่ 2.2



รูปที่ 2.2 การให้พลังงานในแต่ละแบนด์

ขั้นตอนที่ 5 นำค่าพลังงานทั้งหมดมาหาค่าลอการิทึม

ขั้นตอนที่ 6 ทำแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete Cosine Transformation : DCT) ตามสมการ (2.4) ซึ่งก็จะได้ค่าเวกเตอร์สัมประสิทธิ์เซปสตรัมบนสเกลเมล

$$c[i] = \sum_{b=1}^{M} \log_n(E[b]) \times \cos\left(i \times \frac{\pi}{M} \times (b - \frac{1}{2}) \right) \quad (2.4)$$

เมื่อ c คือ เวกเตอร์ของสัมประสิทธิ์เซปสตรัมบนสเกลเมล

i คือ ลำดับของสัมประสิทธิ์เซปสตรัมบนสเกลเมล

M คือ จำนวนแบนด์

b คือ ลำดับแบนด์

2.1.2 การประมาณพันธะเชิงเส้นแบบอิงการรับฟัง (Perceptual Linear Prediction: PLP) [13] และ ราสตา (RelAtive SpecTrAl: RASTA) [14]

ทั้งการประมาณพันธะแบบอิงการรับฟังของมนุษย์และราสตานั้นถูกพัฒนามาจากหลักกการของการประมาณพันธะเชิงเส้น (Linear Predictive Coding : LPC) ให้เหมาะสมกับเสียงพูดของมนุษย์ โดยสัมประสิทธิ์การประมาณพันธะเชิงเส้นนั้นใช้งานอย่างกว้างขวางในด้านการวิเคราะห์เสียง (Speech Analysis) การบีบอัดเสียง(Audio Compression) การจำแนกเสียง ซึ่งอาศัยหลักการพิจารณาค่าต่างๆที่เกิดขึ้นก่อนหน้าแล้วทำนายค่าที่คาดว่าจะเกิดขึ้นโดยอาศัยสมการ (2.5)

$$x[k] = \sum_{i=1}^{n} a_i x[k-i] + \varepsilon[k] \quad (2.5)$$

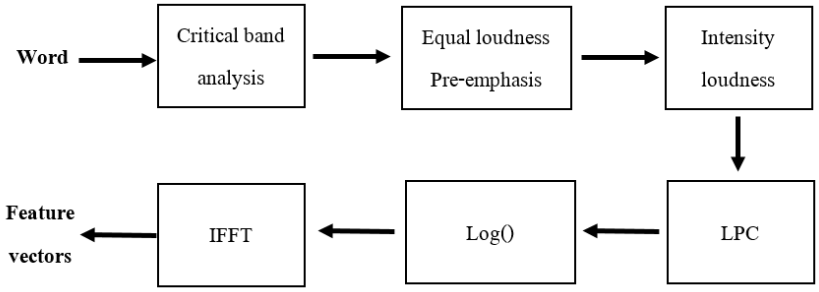
เมื่อ $x[k]$ คือ ค่าปัจจุบันของข้อมูล

$x[k-i]$ คือ ค่าก่อนหน้าของข้อมูล

a_i คือ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น

$\varepsilon[k]$ คือ ค่าผิดพลาดที่คาดว่าจะเกิดขึ้น

โดยการประมาณพันธะแบบอิงการรับฟังของมนุษย์มีการปรับปรุงในส่วนของลักษณะสเปคทรัล (Spectral Characteristic) ให้มีความใกล้เคียงกับระบบการได้ยินของมนุษย์ให้มากขึ้น โดยมีขั้นตอนการทำงานตามรูปที่ 2.3



รูปที่ 2.3 ขั้นตอนการทำงานของการประมาณพันธะแบบอิงการรับฟังของมนุษย์

ขั้นตอนการทำงานเริ่มจากการรับสัญญาณเสียงเข้ามาแล้วทำการแบ่งนับ (Quantization) สัญญาณและแบ่งสัญญาณโดยใช้หน้าต่างแฮมมิ่ง จากนั้นทำการแปลงด้วยกระบวนการแปลงฟูริเยร์แบบเร็ว แล้วนำข้อมูลมาผ่าน ขั้นตอนต่างๆ ดังนี้

การวิเคราะห์คริติคอลแบนด์ (Critical Band Analysis) โดยใช้สมการ (2.6) ซึ่งค่าคริติคอลแบนด์นี้คือการแปลงความถี่ปกติให้อยู่บนสเกลบาร์ก (Bark Scale)

$$\Pi(\alpha) = 6\ln\left[\frac{\alpha}{1200\pi} + \left[\left(\frac{\alpha}{1200\pi}\right)^2 + 1\right]^{0.5}\right] \qquad (2.6)$$

เมื่อ $\Pi(\alpha)$ คือ ค่าคริติคอลแบนด์

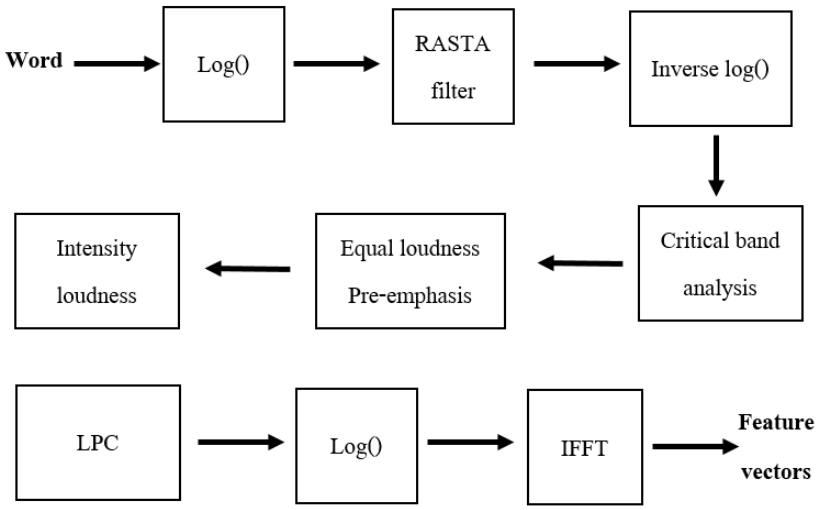
การคำนวณ Intensity loudness root ($I(\alpha)$) จาก Equal loudness ($E(\alpha)$)โดยใช้สมการ (2.7) ซึ่งค่า Equal loudness หาได้จากกราฟ Equal loudness curve

$$I(\alpha) = \left[\Pi(\alpha) \times E(a)\right]^{0.3333} \qquad (2.7)$$

เมื่อ $E(\alpha)$ คือ การประมาณค่าความไวในการได้ยินความถี่ที่แตกต่างกันของมนุษย์

เมื่อผ่านกระบวนการข้างต้นจะได้ค่า Critical Band Filter Output ออกมา จากนั้นนำค่าที่ได้ไปใช้การประมาณพันธะเชิงเส้น หาค่าลอการิทึม และการแปลงฟูริเยร์แบบเร็วย้อนกลับ (Inverse Fast Fourier Transform : IFFT) เพื่อให้ได้ค่าเวกเตอร์ของสัมประสิทธิ์เซปสตรัม (Cepstral Coefficients) ออกมา

ส่วนราสตานั้นเป็นการนำเอาการประมาณพันธะแบบอิงการรับฟังของมนุษย์มาปรับปรุงแก้ไขโดยประยุกต์ใช้ แบนด์พาสฟิลเตอร์ (Band-pass Filter) กับร่วมกับการหาค่าพลังงาน (energy) ในแต่ละความถี่ย่อย (Frequency Sub-band) รวมถึงการแก้ไขทำให้สเปกตรัมระยะสั้น (Short-term Spectrum) มีความนุ่มนวล (Smooth) ยิ่งขึ้น โดยขั้นตอนการทำงานที่เพิ่มขึ้นมาจากประมาณพันธะแบบอิงการรับฟังของมนุษย์คือการนำสเปกตรัมของเสียงหาค่าลอการิทึมและการใช้ตัวกรองราสตา



รูปที่ 2.4 ขั้นตอนการทำงานของราสตา

2.2 การทำเหมืองข้อมูล (Data Mining) [2]

การทำเหมืองข้อมูลเป็นเทคนิคในการนำข้อมูลจำนวนมากมาค้นหารูปแบบความสัมพันธ์ในชุดข้อมูลนั้นๆ เพื่อนำไปใช้ประโยชน์ในงานด้านต่างๆเช่น การวิเคราะห์แนวโน้มของหุ้น, การจำแนกเสียงพูดว่ามีความหมายว่าอะไร, การหาความสัมพันธ์ของการซื้อสินค้าว่า หากซื้อสินค้า A แล้วมีโอกาสที่จะซื้อสินค้า B มากน้อยเพียงใด, การวิเคราะห์พฤติกรรมการเข้าชมเว็ปไซต์ว่าจะมีการเข้าชมก่อนหลังอย่างไร เป็นต้น

โดยก่อนจะนำข้อมูลไปทำเหมืองข้อมูลได้นั้นจำเป็นต้องใช้กระบวนการนอร์มัลไลเซชันข้อมูล (Data Normalization) เพื่อปรับปรุงข้อมูลในส่วนยังผิดพลาดให้เหมาะสมก่อน

ในการทำเหมืองข้อมูลนั้นมีรูปแบบการทำงานอยู่สามประเภทหลักๆคือ

1) กฎความสัมพันธ์ (Association Rule) เป็นการหาความสัมพันธ์ระหว่างข้อมูลโดยอาศัยสองปัจจัยคือ Support Factor เป็นค่าที่บ่งบอกว่าเหตุการณ์ A กับ B มีความถี่ในการเกิดขึ้นมากน้อยแค่ไหน $A \longrightarrow B$: Support Factor = (A U B) และ Confident Factor เป็นค่าที่บอกว่า เมื่อเกิดเหตุการณ์ B แล้ว มีโอกาสที่จะเกิดเหตุการณ์ A มากน้อยแค่ไหน $A \longrightarrow B$: Confident Factor = P(B|A)

2) การจำแนก (Classification) เป็นการสร้างโมเดลเพื่อทำนายกลุ่มข้อมูลที่ไม่เคยเห็นมาก่อนโดยอาศัยเทคนิคในการหาความสัมพันธ์กันในข้อมูล

3) การจับกลุ่ม (Clustering) คือการรวมข้อมูลต่างๆที่มีลักษณะคล้ายกันอยู่ในคลัสเตอร์ (Clusters) เดียวกัน ซึ่งความแตกต่างของการจับกลุ่มและการจำแนกคือ การจับกลุ่มจะไม่ต้องกำหนดคลาสล่วงหน้า และไม่ใช้ข้อมูลในการเรียนรู้ ข้อมูลจะรวมตัวกันบนพื้นฐานของความคล้ายในตัวเอง

ในการทำวิทยานิพนธ์นี้ใช้เทคนิคในด้านการจำแนกในการแก้ปัญหา จากการศึกษาพบว่าเทคนิคในการจำแนกที่ได้รับความนิยมมีมากมายเช่น โครงข่ายประสาทเทียม (Artificial Neural Networks), ฟัซซีโลจิก (Fuzzy Logic), ต้นไม้ตัดสินใจ (Decision Tree), นาอีฟเบย์ (Naïve Bayes), ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine), เกาส์เซียนมิกซ์เจอร์โมเดล (Gaussian Mixture Model) โดยแต่ละเทตนิคนั้นมีข้อดี ข้อเสียแตกต่างกันออกไป โดยในที่นี้จะทำการปรับปรุงโครงข่ายประสาทเทียมร่วมกับฟัซซีโลจิกเพื่อให้ได้ผลการจำแนกที่ดีที่สุด

2.2.1 การนอร์มัลไลเซชันข้อมูล (Data Normalization)

การนอร์มัลไลเซชันข้อมูลเป็นเทคนิคในการทำเหมืองข้อมูลคือการแก้ไขปรับปรุงข้อมูลให้เหมาะสมกับเทคนิคที่ต้องการใช้ ก่อนนำข้อมูลไปใช้ สาเหตุที่ต้องทำการนอร์มัลไลเซชันข้อมูลเพราะข้อมูลอาจยังมีความผิดพลาด ใน 3 ด้านคือ

1. ข้อมูลไม่สมบูรณ์ (Incomplete Data) หรือเกิดค่าของข้อมูลสูญหาย (Missing Value) ซึ่ง เช่น การลืมกรอกข้อมูล

2. ข้อมูลรบกวน (Noisy Data) คือข้อมูลที่มีความผิดพลาดของค่าของข้อมูลไม่ตรงกับแอททริบิวต์ (Attribute) ของข้อมูลเช่น มีเงินเดือน -100 บาท, อายุ -22 ปี, เกิดที่ประเทศเชียงใหม่

3. การไม่สอดคล้องของข้อมูล (Data Inconsistent) คือข้อมูลที่มีความไม่สอดคล้องกันหรือแตกต่างกันเองภายในข้อมูลนั้นๆเช่น ช่วงแรกของข้อมูลเกรดใช้ A B+ B C แต่ช่วงหลังของข้อมูลใช้ 4 3.5 3 2

สาเหตุที่ข้อมูลมีความผิดพลาดอาจเกิดจากปัจจัยต่างๆ ไม่ว่าจะเป็น การรวบรวมข้อมูลจากหลายแหล่ง ความผิดพลาดจากการรับ-ส่งข้อมูล ความผิดพลาดของมนุษย์หรือคอมพิวเตอร์ ดังนั้นจึงมีเทคนิคต่างๆในการปรับปรุงแก้ไขข้อมูลให้พร้อมต่อการนำไปใช้ ซึ่งในที่นี้จะนำเสนอ 3 เทคนิค

2.2.1.1 การนอร์มัลไลเซชันจากค่าสูงต่ำ (Min-Max Normalization)

วิธีนี้จะพิจารณาค่าสูงสุดและค่าต่ำสุดของข้อมูลเดิม ซึ่งจะทำการปรับข้อมูลเดิมเป็นข้อมูลใหม่ โดยให้อยู่ในช่วงที่ต้องการจากค่าสูงสุดและต่ำสุดของข้อมูลใหม่ ซึ่งอาศัยสมการ (2.8)

$$x_i' = \frac{x_i - \min_o}{\max_o - \min_o}(\max_n - \min_n) + \min_n \qquad (2.8)$$

เมื่อ x_i คือข้อมูลเดิมตำแหน่งที่ i

x'_i คือข้อมูลใหม่ตำแหน่งที่ i

$\min_o$ คือค่าต่ำสุดข้อมูลเดิม

$\max_o$ คือค่าสูงสุดข้อมูลเดิม

$\min_n$ คือค่าต่ำสุดข้อมูลใหม่

$\max_n$ คือค่าสูงสุดข้อมูลใหม่

2.2.1.2 การนอร์มัลไลเซชันโดยใช้คะแนนมาตรฐาน (Z-score Normalization)

วิธีการนี้จะพิจารณาปรับข้อมูลจากค่าเฉลี่ย (μ) และค่าเบี่ยงเบนมาตรฐาน (σ) ของข้อมูลเดิม โดยอาศัยสมการ (2.9)

$$x_i' = \frac{x_i - \mu}{\sigma} \qquad (2.9)$$

เมื่อ x_i คือข้อมูลเดิมตำแหน่งที่ i

x'_i คือข้อมูลใหม่ตำแหน่งที่ i

2.2.1.3 การนอร์มัลไลเซชันโดยการปรับจุดทศนิยม (Decimal Normalization)

วิธีการนี้จำขยายช่วงข้อมูลทำให้เห็นจุดทศนิยมหลายตำแหน่งมากขึ้น หรือลดจำนวนจุดทศนิยมเพื่อให้ข้อมูลที่มีจุดทศนิยมหลายตำแหน่งมากๆลดจำนวนลง โดยใช้สมการ (2.10)

$$x_i' = \frac{x_i}{10^j} \tag{2.10}$$

เมื่อ x_i คือข้อมูลเดิมตำแหน่งที่ i

x'_i คือข้อมูลใหม่ตำแหน่งที่ i

j คือจำนวนเต็มบวก

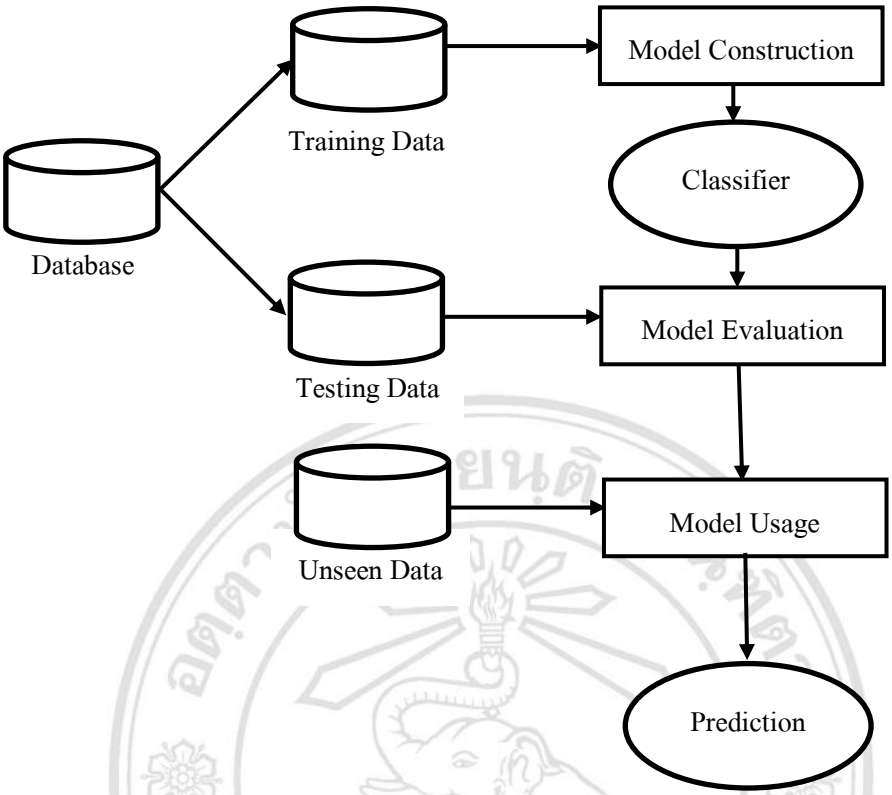
2.2.2 การจำแนกข้อมูล (Classification)

การจำแนกข้อมูลเป็นกระบวนการในการสร้างโมเดลเพื่อระบุคลาสของข้อมูลที่ไม่เคยเห็นมาก่อน (Unseen Data) โดยอาศัยกระบวนการหาความสัมพันธ์ของข้อมูล ซึ่งในกระบวนการนี้จะนำข้อมูลทั้งหมดแบ่งออกเป็นสองกลุ่มคือ ข้อมูลที่ใช้เรียนรู้ (Training Set) ซึ่งเป็นข้อมูลที่ใช้ในการสร้างโมเดลในการจำแนกขึ้นมา และข้อมูลที่ใช้ทดสอบ (Testing Set) เป็นข้อมูลที่ใช้ในการทดสอบประสิทธิภาพของโมเดลที่สร้างขึ้นมาว่ามีความถูกต้องแม่นยำมากน้อยเพียงใด

จากรูที่ 2.5 แสดงกระบวนการสร้างโมเดลจำแนกข้อมูลโดยเริ่มจากนำข้อมูลมาแบ่งเป็นสองกลุ่มคือ ข้อมูลที่ใช้เรียนรู้และข้อมูลที่ใช้ทดสอบ จากนั้นนำข้อมูลที่ใช้เรียนรู้ซึ่งมีการระบุคลาสมาสร้างโมเดล (Model Construction) โดยอาศัยเทคนิคทางเหมืองข้อมูลต่างๆ

เมื่อทำการสร้างโมเดลในการจำแนก (Classifier Model) เสร็จก็จะมาถึงส่วนของการวัดประสิทธิภาพของโมเดล (Model Evaluation) จะเป็นการนำข้อมูลที่ใช้ทดสอบมาทดสอบกับโมเดลการจำแนกที่สร้างขึ้นเพื่อบอกว่าโมเดลนั้นๆมีประสิทธิภาพในการจำแนกมากน้อยเพียงใด แล้วทำการปรับปรุงโมเดลจนกว่าจะมีประสิทธิภาพตามที่คาดหวังไว้

ในส่วนสุดท้ายคือการใช้โมเดล (Model Usage) เป็นขั้นตอนการนำโมเดลการจำแนกที่สร้างขึ้นมาใช้กับข้อมูลที่ไม่เคยเห็นมาก่อนเพื่อทำนายและกำหนดกลุ่มให้กับข้อมูลนั้น



รูปที่ 2.5 กระบวนการสร้างโมเดลจำแนกข้อมูล

2.2.2.1 โครงข่ายประสาทเทียม (Artificial Neural Networks) [15]

โครงข่ายประสาทเทียมเป็นโมเดลทางคณิตศาสตร์ที่จำลองมาจากการทำงานของระบบสมองของมนุษย์คือมีความสามารถการเรียนรู้ ปรับปรุงตัวเองต่อข้อผิดพลาดที่เกิดขึ้น นอกจากนั้นโครงข่ายประสาทเทียมยังมีการหลักทำงานคล้ายระบบประสาทในสมอง คือมีการทำงานร่วมกันในแต่ละโหนด (Node) และส่งผ่านข้อมูลไปมาระหว่างกัน

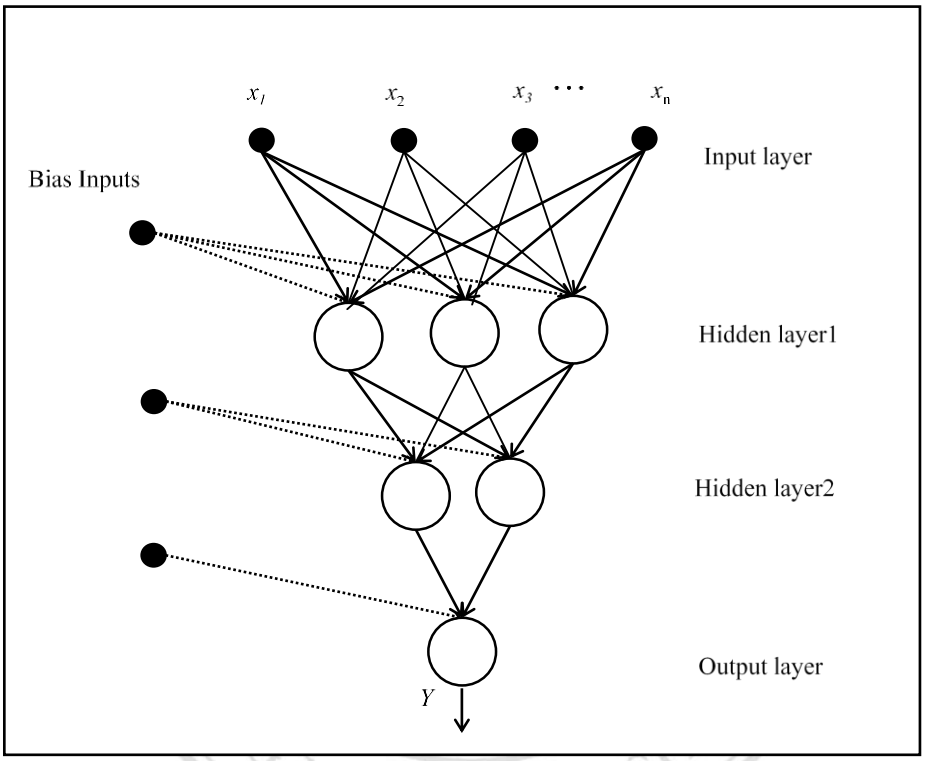
จากการที่โครงข่ายประสาทเทียมมีความฉลาดในการคำนวณ สามารถเรียนรู้ได้ ทนทานต่อข้อผิดพลาด ทำให้เป็นที่ได้รับความนิยมในการนำมาใช้ประโยชน์มากมายในหลายๆ ด้านคือ

1. การจำแนกรูปแบบ (Pattern Recognition) เช่น จำแนกเสียงพูดเพื่อบอกความหาย

2. การทำนาย (Prediction) เช่น ทำนายราคาหุ้น ราคาสินค้า เกรดเฉลี่ย

3. การควบคุม (Control) เช่น การควบคุมหุ่นยนต์ การควบคุมเครื่องจักรในโรงงาน

4. การหาความเหมาะสม (Optimization) เช่น หาระยะทางที่สั้นที่สุดในการเดินทาง (shortest path)



รูปที่ 2.6 ตัวอย่างโครงสร้างโครงข่ายประสาทเทียม

เมื่อพิจารณาในโครงข่ายหนึ่งๆจะประกอบไปด้วยชั้น (Layer) 2 ส่วน คือ ชั้นอินพุต (Input Layer) ชั้นเอาต์พุต (Output Layer) ส่วนชั้นซ่อน (Hidden Layer) จะมีหรือไม่มีก็ได้ หากมีก็จะมีกี่ชั้นซ่อนก็ได้

ขั้นตอนการทำงานของโครงข่ายประสาทเทียมคือชั้นอินพุตจะรับชุดข้อมูล (Data Set) เข้ามา หลังจากนั้นข้อมูลจะถูกไปประมวลผลตามชั้นซ่อน โดยแต่ละชั้นซ่อนประกอบไปด้วยโหนดหลายๆโหนดรวมกันเพื่อช่วยกันประมวลผลโดยมีไบแอส (Bias) ประกอบในแต่ละโหนดโดยถือว่าเป็นอินพุต (Input) 1 ตัว และระหว่างโหนดจะมีค่าถ่วงน้ำหนัก (Weight) เป็นตัวคอยกำหนดความสำคัญของข้อมูลระหว่างโหนด 2 โหนด

ข้อมูลจะเข้ามาผ่านแต่ละชั้นซ่อน ไปตามโหนดออกมาเป็นชั้นเอาต์พุต ซึ่งก็คือคำตอบจากการประมวลผลทั้งเครือข่าย (Network) เหมือนกับการส่งกระแสประสาทไปตามเส้นประสาท โดยจำนวนโหนดของชั้นซ่อน และชั้นเอาต์พุตขึ้นอยู่กับความเหมาะสมของชุดข้อมูลดังรูปที่ 2.6

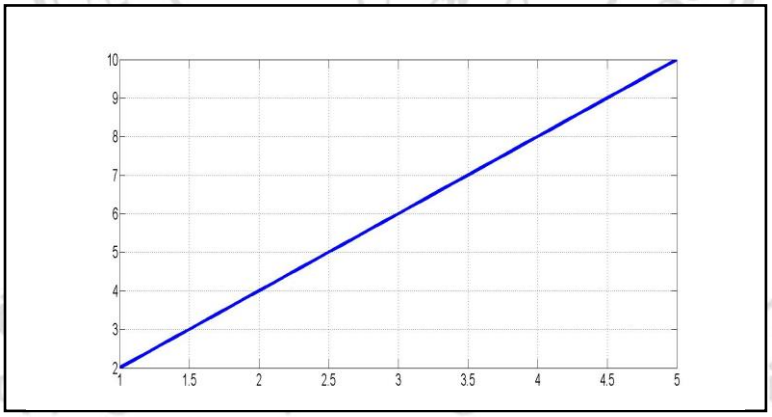
1) ฟังก์ชันกระตุ้น (Activation Function)

ฟังก์ชันกระตุ้นนั้นมีหน้าที่ในการแปลงข้อมูลที่เข้ามาเป็นข้อมูลใหม่ ซึ่งข้อมูลที่เข้ามานั้นก็คือผลรวมของค่าถ่วงน้ำหนักที่เข้ามายังโหนดนั้นๆ โดยอาศัยฟังก์ชันรูปแบบต่าง ซึ่งข้อมูลใหม่ที่ได้นั้นจะอยู่ในช่วงที่เราต้องการ ฟังก์ชันกระตุ้นมีอยู่หลายฟังก์ชันดังนี้

ฟังก์ชันเชิงเส้น (Linear Function)

ฟังก์ชันเชิงเส้นจะมีลักษณะเป็นเส้นตรงโดยมีความชันเท่ากับ n มีลักษณะตามรูปที่ 2.7 โดยใช้สมการตามสมการ (2.11) ซึ่งค่าที่ได้จากฟังก์ชันนี้จะเป็นจำนวนจริง

$$f(v) = nv \tag{2.11}$$



รูปที่ 2.7 ฟังก์ชันเชิงเส้น

ฟังก์ชันขั้น (Step Function)

ฟังก์ชันขั้นเป็นฟังก์ชันที่ให้ค่าเพียง 2 ค่าคือ y_1 และ y_2 โดยอาศัยตัวแบ่ง (α) โดยใช้สมการตามสมการ (2.12) ซึ่งฟังก์ชันขั้นนั้นก็มีหลายรูปแบบเช่น

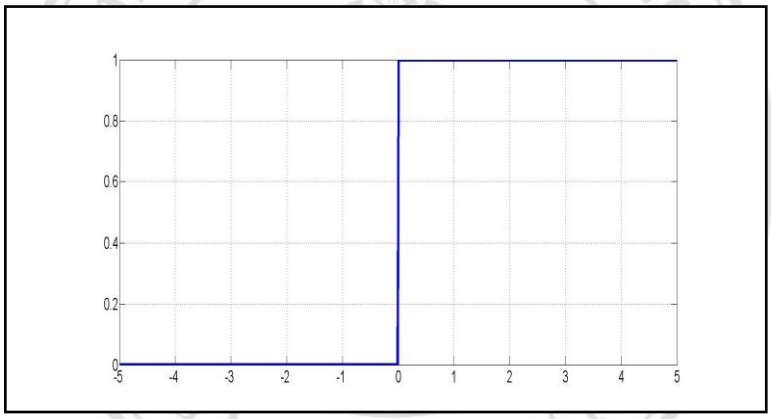
$$f(v) = \begin{cases} y_1; v \geq \alpha \\ y_2; v < \alpha \end{cases} \tag{2.12}$$

ฟังก์ชันจำกัดแข็ง (Hard Limit Function) มีตัวแบ่งที่ 0 และให้ค่าจากฟังก์ชันคือ 0 และ 1 มีสมการตามสมการ (2.13) และลักษณะตามรูปที่ 2.8

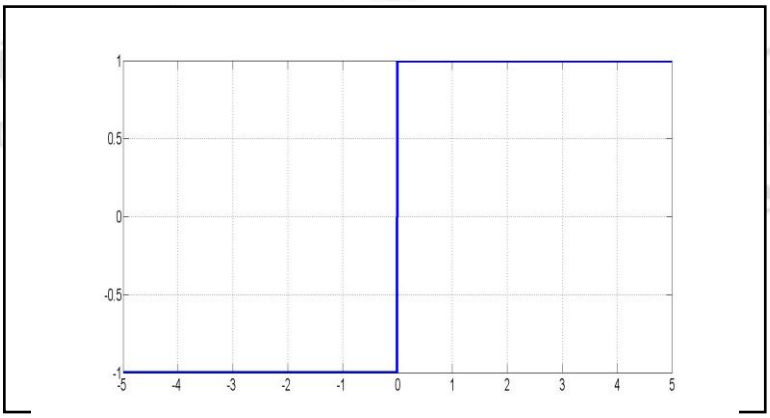
$$f(v)=\begin{cases}1; & v \geq 0\\ 0; & v < 0\end{cases} \tag{2.13}$$

ฟังก์ชันจำกัดแข็งสมมาตร (Symmetric Hard Limit Function) มีลักษณะคล้ายฟังก์ชันแข็ง แต่ให้ค่าจากฟังก์ชันคือ 1 และ -1 มีสมการตามสมการ (2.14) และลักษณะตามรูปที่ 2.9

$$f(v)=\begin{cases}1; & v \geq 0\\ -1; & v < 0\end{cases} \tag{2.14}$$



รูปที่ 2.8 ฟังก์ชันจำกัดแข็ง

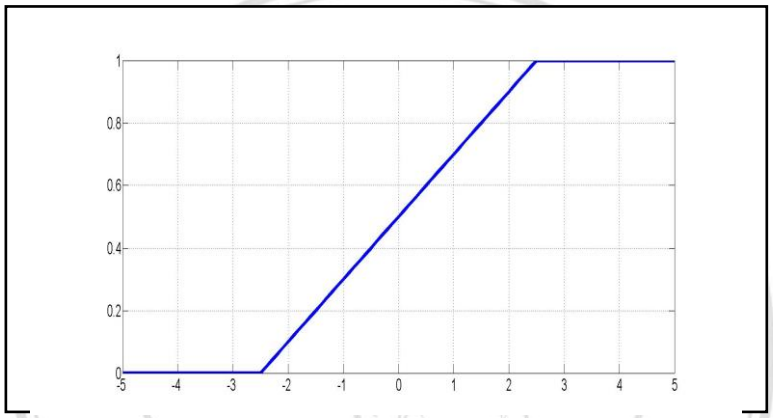


รูปที่ 2.9 ฟังก์ชันจำกัดแข็งสมมาตร

ฟังก์ชันเชิงเส้นจำกัด (Ramp Function)

ฟังก์ชันเชิงเส้นจำกัดเป็นการนำฟังก์ชันเชิงเส้นและฟังก์ชันขั้นมารวมกัน โดยใช้ตัวแบ่งที่ α และ $-\alpha$ มีลักษณะตามรูปที่ 2.10 และใช้สมการ (2.15)

$$f(v) = \begin{cases} y; v \geq a \\ v; -a < v < a \\ -y; v \leq -a \end{cases} \tag{2.15}$$



รูปที่ 2.10 ฟังก์ชันเชิงเส้นจำกัด

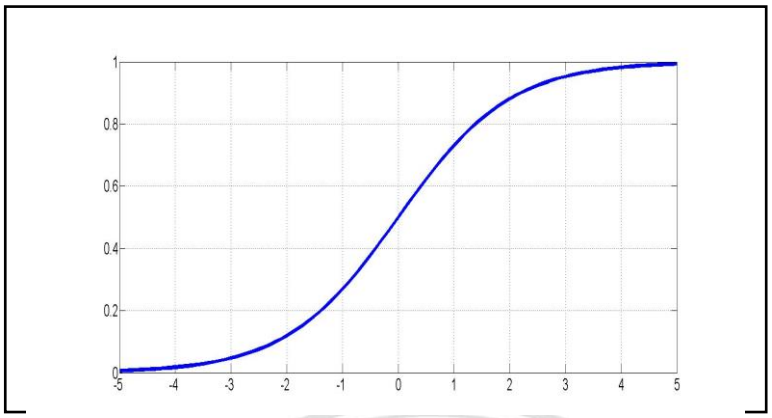
ฟังก์ชันซิกมอยด์ (Sigmoid Function)

ฟังก์ชันซิกมอยด์เป็นฟังก์ชันที่มีลักษณะฟังก์ชันคล้ายรูปตัว S เป็นฟังก์ชันได้รับความนิยมเป็นอย่างมาก เพราะค่าที่ได้จากฟังก์ชันนั้นจะค่อยๆเพิ่มขึ้นทีละน้อย ฟังก์ชันซิกมอยด์มีหลายรูปแบบเช่น

ฟังก์ชันลอกซิกมอยด์ (Logarithmic Sigmoid function) เป็นฟังก์ชันตามรูปที่ 2.11 โดยมีสมการตามสมการ (2.16)

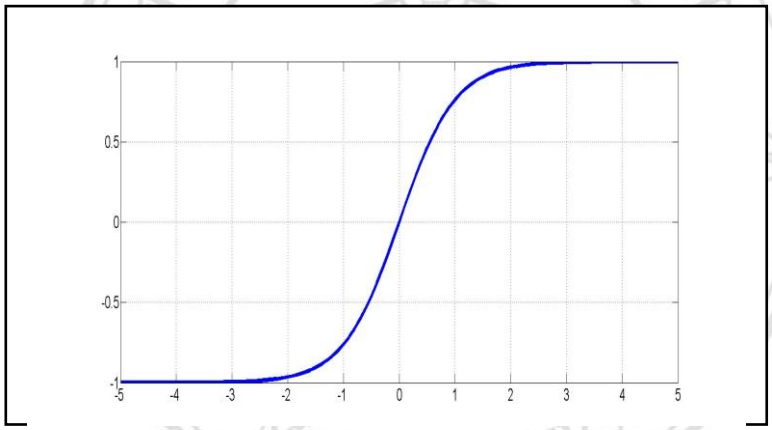
$$f(v) = \frac{1}{1+e^{-v}} \tag{2.16}$$

ฟังก์ชันไฮเปอร์บอลิกแทนเจนต์ซิกมอยด์ (Hyperbolic Tangent Sigmoid Function) เป็นฟังก์ชันตามรูปที่ 2.12 โดยมีสมการตามสมการ (2.17)



รูปที่ 2.11 ฟังก์ชันลอกซิกมอยด์

$$f(v) = \frac{2}{1+e^{-2v}} - 1 \tag{2.17}$$



รูปที่ 2.12 ฟังก์ชันไฮเปอร์บอลิกแทนเจนต์ซิกมอยด์

ฟังก์ชันเกาส์เซียน (Gaussian Function)

ฟังก์ชันเกาส์เซียนเป็นฟังก์ ชันที่อาศัยค่าเฉลี่ย (μ) หาได้จากสมการ (2.18) และค่าเบี่ยงเบนมาตรฐาน (σ) หาได้จากสมการ (2.19) ฟังก์ชันเกาส์เซียนมีลักษณะตามรูปที่ 2.13 โดยมีสมการตามสมการ (2.20)

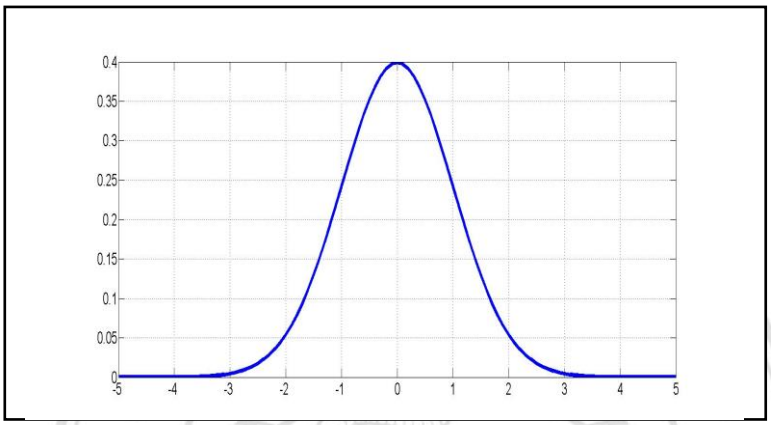
$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.18}$$

เมื่อ n คือจำนวนข้อมูลทั้งหมด

x_i คือข้อมูลตำแหน่งที่ i

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}} \tag{2.19}$$

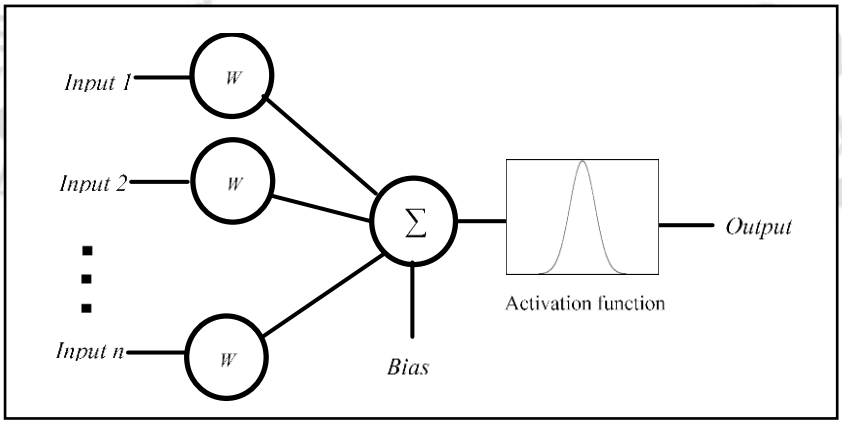
$$f(v) = e^{-\left(\frac{v-\mu}{\sigma}\right)^2} \tag{2.20}$$



รูปที่ 2.13 ฟังก์ชันเกาส์เซียน

2) เปอร์เซปทรอนหลายชั้น (Multi-Layer Perceptrons : MLP)

ส่วนประกอบของเปอร์เซปทรอนหนึ่งชั้นจากรูปที่ 2.14 นั้นจะประกอบไปด้วยอินพุตที่รับเข้ามาหรือคือค่าคุณลักษณะ, ค่าถ่วงน้ำหนักของอินพุตเหล่านั้นซึ่งเกิดจากการสุ่ม, ค่าไบแอสเปรียบเสมือนอินพุตหนึ่งๆ ส่วนฟังก์ชันกระตุ้นซึ่งเป็นฟังก์ชันที่ใช้ในการแบ่งค่าที่เข้ามา, เอาต์พุต (Output) คือค่าที่ได้จากการผ่านฟังก์ชันกระตุ้น



รูปที่ 2.14 แสดงส่วนประกอบต่างๆของเปอร์เซปทรอน 1 โหนด

โดยก่อนจะไปคำนวณส่วนฟังก์ชันกระตุ้นนั้น จะทำการหาผลรวมค่าถ่วงน้ำหนักของอินพุตทุกตัวจากสมการ (2.21)

$$v = \sum_{i=1}^{n} w_i x_i + Bias \qquad (2.21)$$

เมื่อ v คือผลรวมของค่าถ่วงน้ำหนัก

x คือค่าอินพุต

w คือค่าถ่วงน้ำหนัก

จากนั้นนำค่า v จากสมการ (2.21) ไปผ่านฟังก์ชันกระตุ้นเป็นค่า v'

ในการใช้งานจริงๆนั้นเปอร์เซปทรอน ชั้นเดียวย่อมมีประสิทธิภาพในการคำนวณน้อย จึงมีการนำ เปอร์เซปทรอนหลายๆอันมารวมกันเป็นเปอร์เซปทรอนหลายชั้น โดยกระบวนการทั้งหมดจนได้ค่าคำตอบออกมา เรียกว่า โครงข่ายป้อนค่าไปข้างหน้า (Feed Forward)

3) การแพร่กระจายย้อนกลับ (Backpropagation)

จุดเด่นในการทำงานของโครงข่ายประสาทเทียมนั้นคือการเรียนรู้และแก้ไขตัวเองจากความผิดพลาดผิดพลาดที่เกิดขึ้น ซึ่งการเรียนรู้ของโครงขายประสาทเทียมนั้นมี 2 รูปแบบคือ

1. การเรียนรู้แบบมีการสอน (Supervised Learning) เป็นวิธีการเรียนรู้โดยอาศัยการปรับปรุงความผิดพลาดที่เกิดขึ้นจากการเรียนรู้ ดังนั้นการเรียนรู้แบบนี้จำเป็นต้องรู้ถึงคลาสของข้อมูล เพื่อจะสามารถตรวจสอบความผิดพลาดที่เกิดขึ้นในการเรียนรู้แต่ละรอบได้

2. การเรียนรู้แบบไม่มีการสอน (Unsupervised Learning) เป็นวิธีการเรียนรู้ที่จะไม่ทราบถึงคลาสของข้อมูล ทำให้ไม่สามารถตรวจสอบข้อผิดพลาดที่เกิดขึ้นระหว่างการสอนได้ หลักการทำงานของการเรียนรู้แบบไม่มีการสอนคือจะทำการจับกลุ่มข้อมูลที่มีความใกล้เคียงกันไว้เป็นคลาสเดียวกัน

การแพร่กระจายย้อนกลับเป็นหนึ่งในอัลกอริทึมของโครงข่ายประสาทเทียมแบบการเรียนรู้แบบมีการสอน ซึ่งเป็นที่นิยมอย่างมากในการนำมาใช้งาน หลักการทำงานคือเริ่มจากเมื่อโครงข่ายประสาทเทียมเรียนรู้ชุดข้อมูลเสร็จแล้ว ก็จะหาค่าความผิดพลาดที่เกิดขึ้น จากนั้นนำค่าที่ได้มาทำการแก้ไขโครงสร้างในส่วนของค่าถ่วงน้ำหนัก โดยค่าความผิดพลาดนั้นสามารถหาได้จากสมการ (2.22) โดยเอาต์พุตที่ต้องการ (Desirer Output) คือ ค่าเอาต์พุตจริงๆของข้อมูลนั้นๆ

$$e_n = d_n - y_n \qquad (2.22)$$

เมื่อ e คือความผิดพลาด (error) ของแต่ละโหนด

d คือเอาต์พุตที่ต้องการ

y คือเอาต์พุต

n คือตำแหน่งโหนด

ในกระบวนการแพร่กระจายย้อนกลับนั้นจะทำการหาค่า Δw เพื่อนำไปใช้ในการแก้ไขปรับปรุงโครงสร้างของโครงข่ายประสาทเทียมในส่วนของค่าถ่วงน้ำหนักตามสมการ (2.23) โดยกำหนดค่าอัตราการเรียนรู้ (Learning Rate) ขึ้นมาเอง หากค่าอัตราการเรียนรู้มากก็จะทำให้การเปลี่ยนแปลงค่าถ่วงน้ำหนักในแต่ละรอบน้อย

$$\Delta w(n) = -\eta * \delta(n) * x(n) \tag{2.23}$$

เมื่อ η คือค่าอัตราการเรียนรู้

x คืออินพุต

δ คือเกรเดียนต์

โดยค่าเกรเดียนต์ (Gradient) จะแบ่งออกเป็น *2* กรณีคือเกรเดียนต์สำหรับชั้นเอาต์พุต และชั้นซ่อนซึ่งใช้สมการต่างกันตามสมการ (2.24)

$$\delta(n) = \begin{cases} e(n) * f'(v(n)) & ; \text{Output} \\ f'(v(n)) * \sum \delta_i(n) * w_i(n) & ; \text{Hidden} \end{cases} \tag{2.24}$$

เมื่อ f' คือฟังก์ชันกระตุ้นที่ผ่านการหาค่าค่าอนุพันธ์

i คือเกรเดียนต์ของชั้นถัดไป

ซึ่งค่าถ่วงน้ำหนักใหม่หาได้จากสมการ (2.25)

$$w(n+1) = w(n) + \Delta w \tag{2.25}$$

2.2.2.2 ฟัซซีโลจิก (Fuzzy logic) [14]

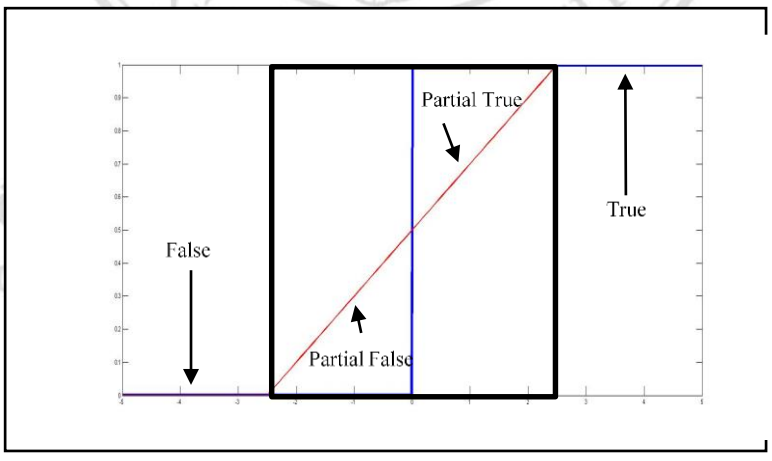
ตรรกศาสตร์คือการให้เหตุผลของสิ่งต่างๆที่เกิดขึ้น ในอดีตนั้นมีเพียงตรรกศาสตร์แบบจริงเท็จ (Boolean Logic)โดยใช้เซทแบบดั่งเดิม (Crisp Set) ซึ่งเซทแบบดั่งเดิมมีขอบเขตที่ชัดเจน มีค่าความเป็นสมาชิกเพียงสองค่าคือ จริง (Completely True) และเท็จ (Completely False) หรือใช่และไม่ใช่ หรือ 0 และ 1 แต่ในชีวิตจริงนั้นเราไม่สามารถบอกสิ่งที่เกิดขึ้นต่างๆ รอบตัวเราด้วยค่าเพียงสอง

ค่าได้ จึงมีการพัฒนาฟัซซีโลจิกมาเพื่อใช้อธิบายเหตุการณ์ต่างๆที่เกิดขึ้นรอบตัว ซึ่งมักจะใช้ในเหตุการณ์ที่เกี่ยวข้องกับความรู้สึกของมนุษย์เช่น

หากมีคนถามว่า วันนี้รู้สึกว่าอากาศเป็นอย่างไร การจะตอบว่าวันนี้ร้อน หนาว อบอุ่นนั้น แต่ละคนย่อมตอบแตกต่างกันไป เช่นถ้าอุณหภูมิ 25 องศาเซลเซียส คนที่อยู่ภาคกลางก็จะบอกว่าอากาศเย็นต้องหาเสื้อผ้ามาใส่เพิ่ม ส่วนคนที่อยู่ภาคเหนือก็จะบอกว่าอากาศกำลังดี

เมื่อถามช่วงน้ำหนักว่าคนจะอ้วนหรือผอม ที่น้ำหนักประมาณเท่าไหร่ คนจะสูงหรือเตี้ยที่ส่วนสูงเท่าไหร่ แต่ละคนย่อมบอกช่วงน้ำหนักและส่วนสูงที่แตกต่างกันออกไป บางคนก็บอกว่าคนอ้วนคือคนที่มีน้ำหนักตั้งแต่ 75 กิโลกรัมขึ้นไป บางคนก็บอก 70 กิโลกรัมขึ้นไป บางคนก็บอกว่าสูง 170 เซนติเมตรถือว่าสูงแล้ว บางคนก็บอกว่าสูง 175 เซนติเมตรถึงจะเรียกว่าสูง ซึ่งมีขอบเขตที่ไม่แน่นอน

จากตัวอย่างข้างต้นจะเห็นได้ว่า ลักษณะของฟัซซีโลจิกคือมีความไม่แน่นอน ดังนั้นจึงใช้เซทฟัซซี (Fuzzy Set) แทนที่จะเป็นเซทแบบดั่งเดิม โดยมีลักษณะคือมีขอบเขตที่ไม่จำกัด (Un-sharp Boundary) และมีค่าความเป็นสมาชิกอยู่ระหว่างจริง ช่วงการต่อขยายความจริง (Partial True) ช่วงการต่อขยายความเท็จ (Partial False) และเท็จตามรูปที่ 2.15 ซึ่งเส้นสีน้ำเงินแทนค่าความเป็นสมาชิกของตรรกศาสตร์แบบดั่งเดิม และเส้นสีแดงแทนค่าความเป็นสมาชิกของฟัซซีโลจิก



รูปที่ 2.15 แสดงค่าความจริงของตรรกศาสตร์ดั่งเดิมและฟัซซีโลจิก

1) ฟังก์ชันความเป็นสมาชิก (Membership Function)

ในการบอกค่าความเป็นสมาชิก (Membership Value : μ) ของฟัซซีโลจิกนั้นจะอาศัยฟังก์ชันความเป็นสมาชิก โดยข้อมูลที่เข้ามาหนึ่งข้อมูลอาจมีค่าความเป็นสมาชิกหลากค่าได้ จึงทำให้ฟังก์ชัน

ความเป็นสมาชิกสามารถซ้อนทับกันได้ ค่าความเป็นสมาชิกนั้นจะอยู่ในช่วง 0 ถึง 1 เท่านั้น โดยฟังก์ชันความเป็นสมาชิกที่ได้รับความนิยมมีหลากหลายฟังก์ชัน ดังนั้นจึงควรเลือกให้เหมาะสมกับงานที่ใช้

ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน (Gaussian Membership Function)

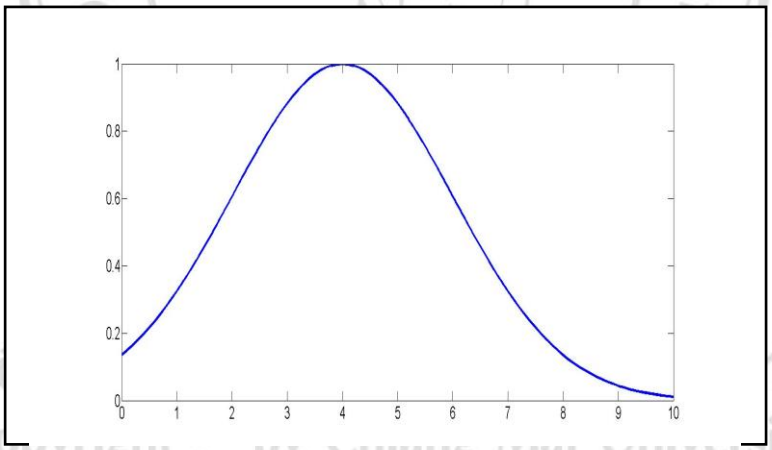
ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียนนั้นอาศัยตัวแปร 2 ตัวคือค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน โดยมีสมการตามสมการ (2.26)

$$\mu_{Gau}(x:\alpha,\sigma) = \exp\left(-\frac{(x-\alpha)^2}{2\sigma^2}\right) \tag{2.26}$$

เมื่อ α คือค่าเฉลี่ย

σ คือค่าเบี่ยงเบนมาตรฐาน

โดยรูปที่ 2.16 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน มีค่าเฉลี่ยเท่ากับ 4 และค่าเบี่ยงเบนมาตรฐานเท่ากับ 2



รูปที่ 2.16 ฟังก์ชันความเป็นสมาชิกแบบเกาส์เซียน

ฟังก์ชันความเป็นสมาชิกแบบระฆังคว่ำ (Bell-shape Membership Function)

ฟังก์ชันความเป็นสมาชิกแบบระฆังคว่ำนั้นอาศัยตัวแปร 3 ตัวคือ a b และ c โดยมีสมการตามสมการ (2.27)

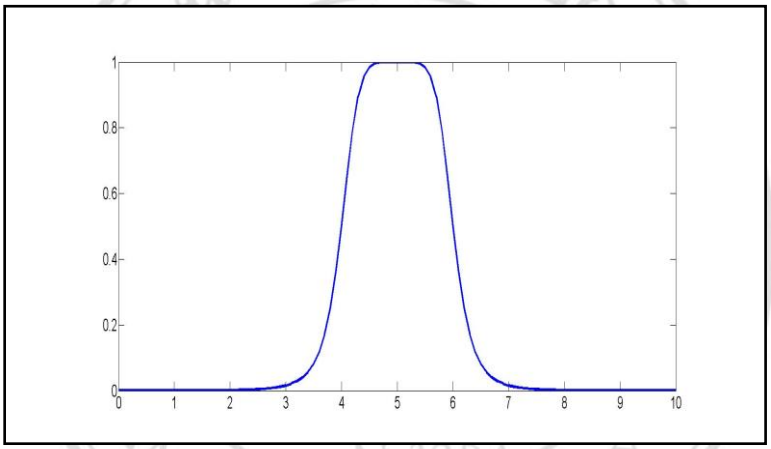
$$\mu_{Bell}(x:a,b,c)=\frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{2.27}$$

เมื่อ a คือความกว้างของกราฟ

b คือค่าที่ใช้ควบคุมความชัน

c คือตำแหน่งกึ่งกลางกราฟ

โดยรูปที่ 2.17 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบระฆังคว่ำ มีค่า a เท่ากับ 1, b เท่ากับ 3 และ c เท่ากับ 5



รูปที่ 2.17 ฟังก์ชันความเป็นสมาชิกแบบระฆังคว่ำ

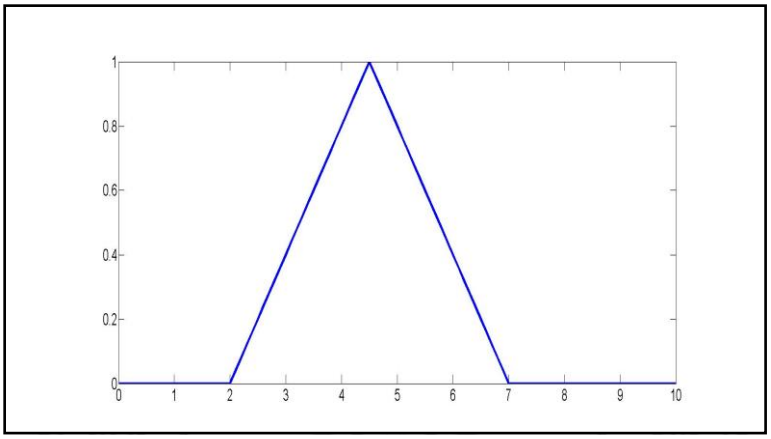
ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม (Triangular Membership Function)

ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยมนั้นอาศัยตัวแปร 3 ตัวคือ a b และ c มีสมการตามสมการ (2.28)

$$\mu_{Tri}(x:a,b,c)=\begin{cases}0; & x<a\\ \frac{x-a}{b-a}; & a\le x<b\\ \frac{c-x}{c-b}; & b\le x\le c\\ 0; & x>c\end{cases} \tag{2.28}$$

เมื่อ a, c คือตำแหน่งของพื้นของรูปสามเหลี่ยม

b คือตำแหน่งของจุดยอด



รูปที่ 2.18 ฟังก์ชันความเป็นสมาชิกรูปสามเหลี่ยม

โดยรูปที่ 2.18 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยม มีค่า a เท่ากับ 2, b เท่ากับ 4.5 และ c เท่ากับ 7

ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู (Trapezoidal Membership Function)

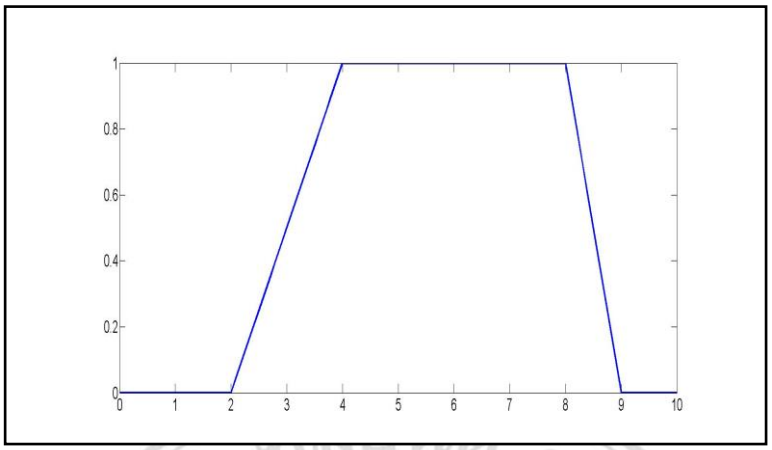
ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมูนั้นอาศัยตัวแปร 4 ตัวคือ a b c และ d มีสมการตามสมการ (2.29)

$$\mu_{Tra}(x:a,b,c,d)=\begin{cases}0;\ x<a\\ \dfrac{x-a}{b-a};\ a\le x<b\\ 1;\ b\le x<c\\ \dfrac{d-x}{d-c};\ c\le x<d\\ 0;\ x\ge d\end{cases} \qquad (2.29)$$

เมื่อ a, d คือตำแหน่งของพื้นของรูปสี่เหลี่ยม

b, c คือตำแหน่งของเส้นคู่ขนานกับพื้นสี่เหลี่ยม

โดยรูปที่ 2.19 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบสี่เหลี่ยมคางหมู มีค่า a เท่ากับ 2, b เท่ากับ 4, c เท่ากับ 7 และ d เท่ากับ 9



รูปที่ 2.19 ฟังก์ชันความเป็นสมาชิกรูปสี่เหลี่ยมคางหมู

ฟังก์ชันความเป็นสมาชิกแบบเชิงเส้น (Linear Membership Function)

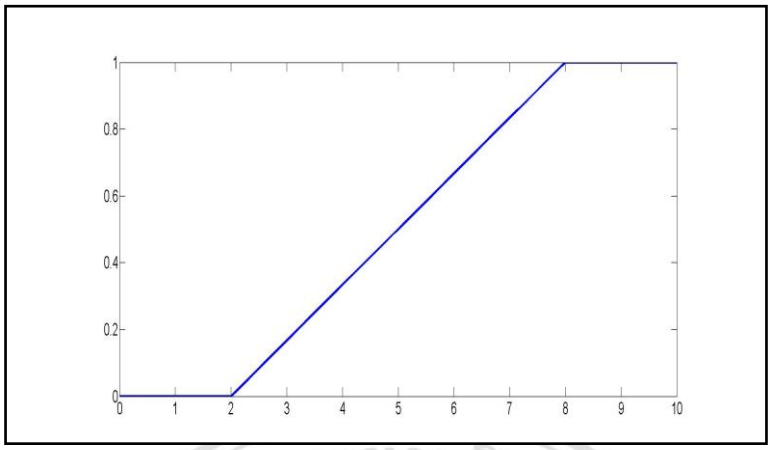
ฟังก์ชันความเป็นสมาชิกแบบเชิงเส้นนั้นอาศัยตัวแปร 2 ตัวคือ a และ b โดยมีรูปแบบของกราฟ 2 รูปแบบคือ เชิงเส้นบวก (Positive linear) มีสมการตามสมการ (2.30) และเชิงเส้นลบ Negative linear) มีสมการตามสมการ (2.31)

$$\mu_{LinearP}(x:a,b)=\begin{cases}0;\ x<a\\ \dfrac{x-a}{b-a};\ a\le x<b\\ 1;\ x\ge b\end{cases} \tag{2.30}$$

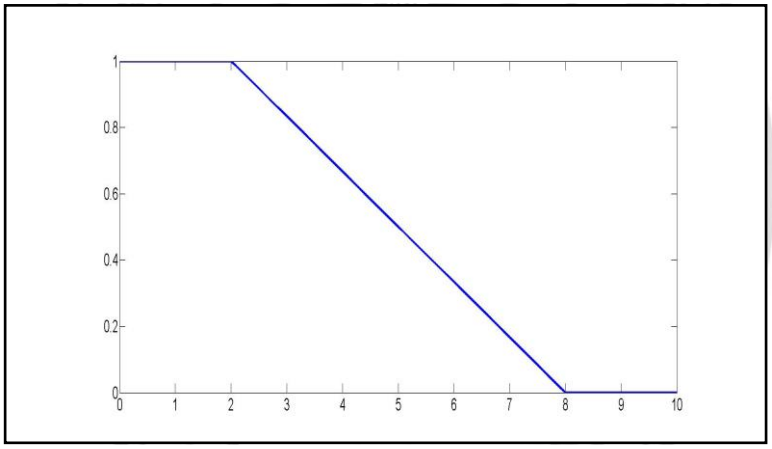
$$\mu_{LinearN}(x:a,b)=\begin{cases}0;\ x<a\\ \dfrac{a-x}{b-a};\ a\le x<b\\ 1;\ x\ge b\end{cases} \tag{2.31}$$

เมื่อ a,b คือตำแหน่งเปลี่ยนความชัน

โดยรูปที่ 2.20 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบเชิงเส้นบวก มีค่า a เท่ากับ 2 และ b เท่ากับ 8 ส่วนรูปที่ 2.21 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบเชิงเส้นลบ มีค่า a เท่ากับ 8 และ b เท่ากับ 2



รูปที่ 2.20 ฟังก์ชันความเป็นสมาชิกแบบเชิงเส้นบวก



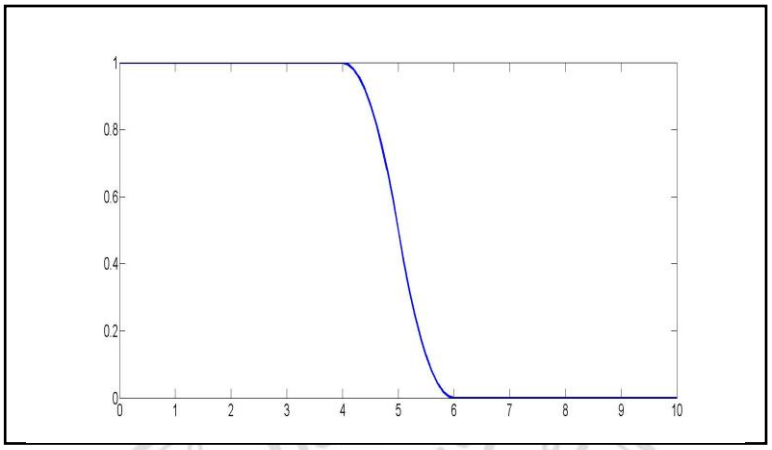
รูปที่ 2.21 ฟังก์ชันความเป็นสมาชิกแบบเชิงเส้นลบ

ฟังก์ชันความเป็นสมาชิกรูปตัวแซด (Z Membership Function)

ฟังก์ชันความเป็นสมาชิกรูปตัวแซดนั้นอาศัยตัวแปร 2 ตัวคือ a และ b มีสมการตามสมการ (2.32)

$$\mu_Z(x:a,b)=\begin{cases}1;\ x<a\\ 1-2\left(\dfrac{x-a}{b-a}\right)^2;\ a\le x<\dfrac{a+b}{2}\\ 2\left(b-\dfrac{x}{b-a}\right)^2;\ \dfrac{a+b}{2}\le x<b\\ 0;\ x\ge b\end{cases} \tag{2.32}$$

เมื่อ a,b คือตำแหน่งเปลี่ยนความชัน



รูปที่ 2.22 ฟังก์ชันความเป็นสมาชิกรูปตัวแซด

โดยรูปที่ 2.22 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกรูปตัวแซด มีค่า a เท่ากับ 4 และ b เท่ากับ 6

ฟังก์ชันความเป็นสมาชิกรูปตัวเอส (S Membership Function)

ฟังก์ชันความเป็นสมาชิกรูปตัวเอสนั้นอาศัยตัวแปร 2 ตัวคือ a และ b มีสมการตามสมการ (2.33)

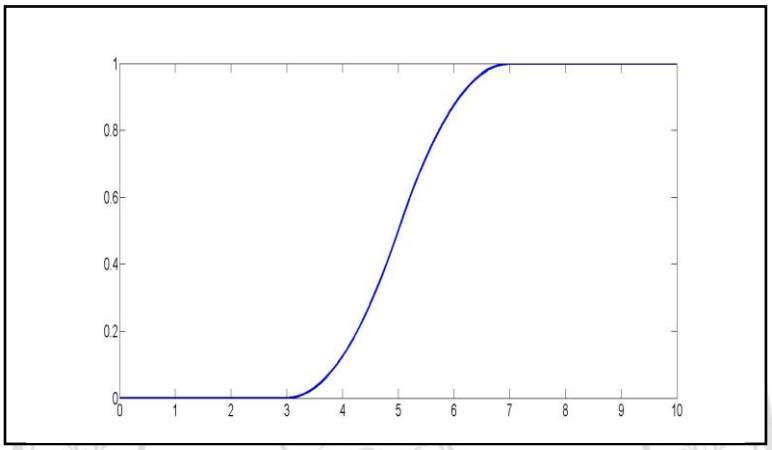
$$\mu_Z(x:a,b)=\begin{cases}1; x<a\\ 2\left(\dfrac{x-a}{b-a}\right)^2; a\le x<\dfrac{a+b}{2}\\ 1-2\left(\dfrac{x-b}{b-a}\right)^2; \dfrac{a+b}{2}\le x<b\\ 0; x\ge b\end{cases} \tag{2.33}$$

เมื่อ a,b คือตำแหน่งเปลี่ยนความชัน

โดยรูปที่ 2.23 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกรูปตัวเอส มีค่า a เท่ากับ 3 และ b เท่ากับ 7

ฟังก์ชันความเป็นสมาชิกแบบซิกมอยด์ (Sigmoidal Membership Function)

ฟังก์ชันความเป็นสมาชิกแบบซิกมอยด์นั้นอาศัยตัวแปร 2 ตัวคือ a และ b โดยหากค่า a มีค่าน้อยจะให้กราฟที่มีความชันต่ำ แต่ถ้าค่า a มากจะให้กราฟที่มีความชันสูง ค่า b คือค่าที่มีค่าความเป็นสมาชิกที่ 0.5 มีสมการตามสมการ (2.34)



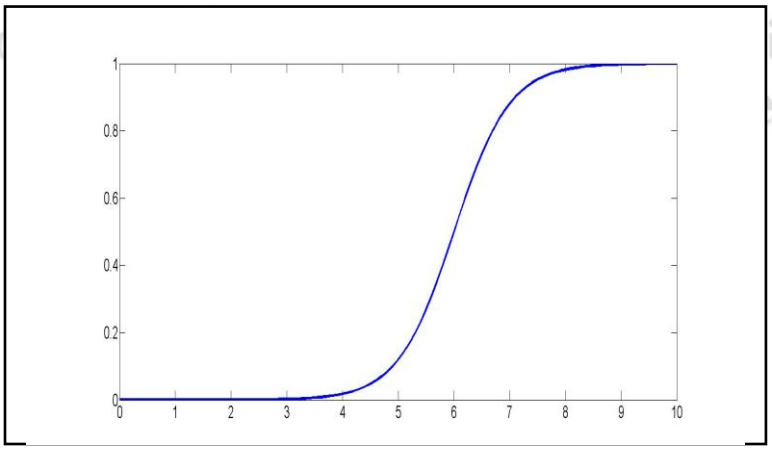
รูปที่ 2.23 ฟังก์ชันความเป็นสมาชิกรูปตัวเอส

$$\mu_{Sig}(x:a,b)=\frac{1}{1+\exp(-a(x-b))} \tag{2.34}$$

เมื่อ a คือค่ากำหนดความชัน

b คือค่าที่ให้ค่าความเป็นสมาชิก 0.5

โดยรูปที่ 2.24 แสดงตัวอย่างของฟังก์ชันความเป็นสมาชิกแบบซิกมอยด์ มีค่า a เท่ากับ 2 และ b เท่ากับ 6



รูปที่ 2.24 ฟังก์ชันความเป็นสมาชิกแบบซิกมอยด์

2) โครงสร้างและขั้นตอนการทำงานของฟัซซีโลจิก

โครงสร้างโดยรวมของฟัซซีโลจิกนั้นเป็นไปตามรูปที่ 2.25 โดยจะประกอบไปด้วย 4 ส่วนหลักๆคือ

1. ส่วนการแปลงข้อมูล ในส่วนนี้จะรับข้อมูลปกติซึ่งเป็นเซทแบบดั่งเดิมมาทำการแปลงเป็นฟัซซีเซทหรือตัวแปรภาษา (Linguistic Variable) กระบวนการการแปลงข้อมูลนี้เรียกว่า Fuzzification ในการแปลงข้อมูลนั้นอาศัยฟังก์ชันความเป็นสมาชิกในการแปลง

2. ส่วนฐานความรู้ (Knowledge Base) มีหน้าที่ในการเก็บและรวบรวมข้อมูลเพื่อใช้ในการควบคุมส่วนอื่นๆของฟัซซีโลจิก โดยประกอบไปด้วย 2 ส่วนคือ ส่วนฐานข้อมูล (Data base) เป็นส่วนที่ใช้เก็บรวบรวมและจัดเตรียมข้อมูลเช่น ฟังก์ชันความเป็นสมาชิกจะใช้ฟังก์ชันใดบ้างในแต่ละค่าคุณลักษณะซึ่งในหนึ่งค่าคุณลักษณะอาจมีหลายฟังก์ชัน รวมไปถึงแต่ละฟังก์ชันใช้ค่าตัวแปรเท่าใด ส่วนฐานกฎ (Rule Base) มีหน้าที่ในการกำหนดกฎ (Rule) ที่จะในไปใช้ โดยกฎนี้ได้มาจากผู้เชี่ยวชาญ ซึ่งจะอยู่ในรูปของชุดข้อมูลแบบกฎเชิงภาษา (Linguistic Rule)

3. ส่วนการตีความ (Inference Engine) มีหน้าที่ในการตรวจสอบความถูกผิด ข้อเท็จจริง รวมทั้งวิธีที่จะใช้เพื่อนำไปใช้ในการหาคำตอบโดยอาศัยการตีความ การตีความนี้มีลักษณะคล้ายกับการใช้ความรู้ในการแก้ไขปัญหาที่เกิดขึ้นได้อย่างถูกต้องเหมาะสม

4. ส่วนการแปลงค่าฟัซซีกลับเป็นค่าทั่วไป (Defuzzification) มีหน้าที่ในการสรุปคำตอบสุดท้าย เพื่อนำคำตอบนี้ไปใช้งานต่อไป โดยอาศัยการรวมกฎ (Aggregation) ที่ได้จากการตีความในขั้นตอนก่อนหน้า โดยวิธีการรวมกฎนั้นก็มีอยู่หลายวิธีเช่น

วิธีหาค่าพื้นที่กลาง (Centroid of Area) หรือวิธีเฉลี่ยถ่วงน้ำหนัก (Weighted Average Method) วิธีนี้จะเป็นหาค่าเฉลี่ยของแต่ละจุดจากพื้นที่ใต้กราฟที่ได้มาจากการตีความโดยอาศัยสมการ (2.35) โดยเรียกค่าที่ได้ว่า ค่าจุดศูนย์ถ่วงโดยรวม (Central of gravity : C)

$$C = \frac{\sum_{i=1}^{n} \mu(y_i) * y_i}{\sum_{i=1}^{n} \mu(y_i)} \qquad (2.35)$$

เมื่อ $\mu(y_i)$ คือค่าความเป็นสมาชิกของข้อมูลตำแหน่งที่ i

y_i คือค่าจริงของข้อมูลตำแหน่งที่ i

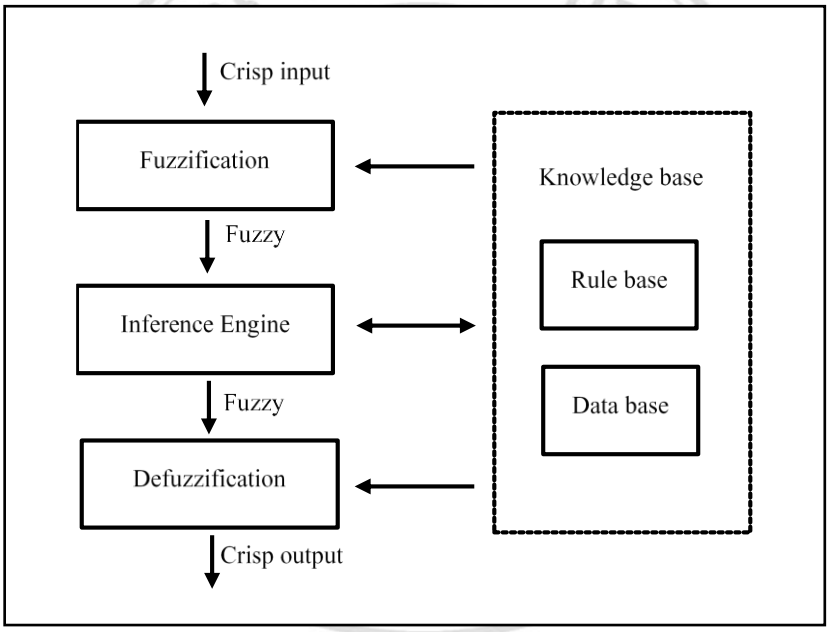
วิธีแบ่งครึ่งพื้นที่ (Bisector of Area : B) วิธีนี้จะได้ค่าคำตอบเป็นครึ่งหนึ่งของพื้นที่ใต้กราฟ สามารหาได้จากสมการ (2.36)

$$B = \left\{ y_a \mid \sum_{i=1}^{a} \mu(y_i) \geq \frac{\sum_{i=1}^{n} \mu(y_i)}{2} \right\} \tag{2.36}$$

เมื่อ $\mu(y_i)$ คือค่าความเป็นสมาชิกของข้อมูลตำแหน่งที่ i

y_i คือค่าจริงของข้อมูลตำแหน่งที่ i

a คือตำแหน่งที่แบ่งครึ่งพื้นที่ใต้กราฟ



รูปที่ 2.25 โครงสร้างของตรรกศาสตร์คลุมเครือ

วิธีค่าเฉลี่ยของค่าสูงสุด (Mean Value of Maximum : M) วิธีนี้จะได้คำตอบจากการคำนวณค่าเฉลี่ยที่มีดรีกรีความเป็นสมาชิกสูงสุด โดยใช้สมาการ (2.37)

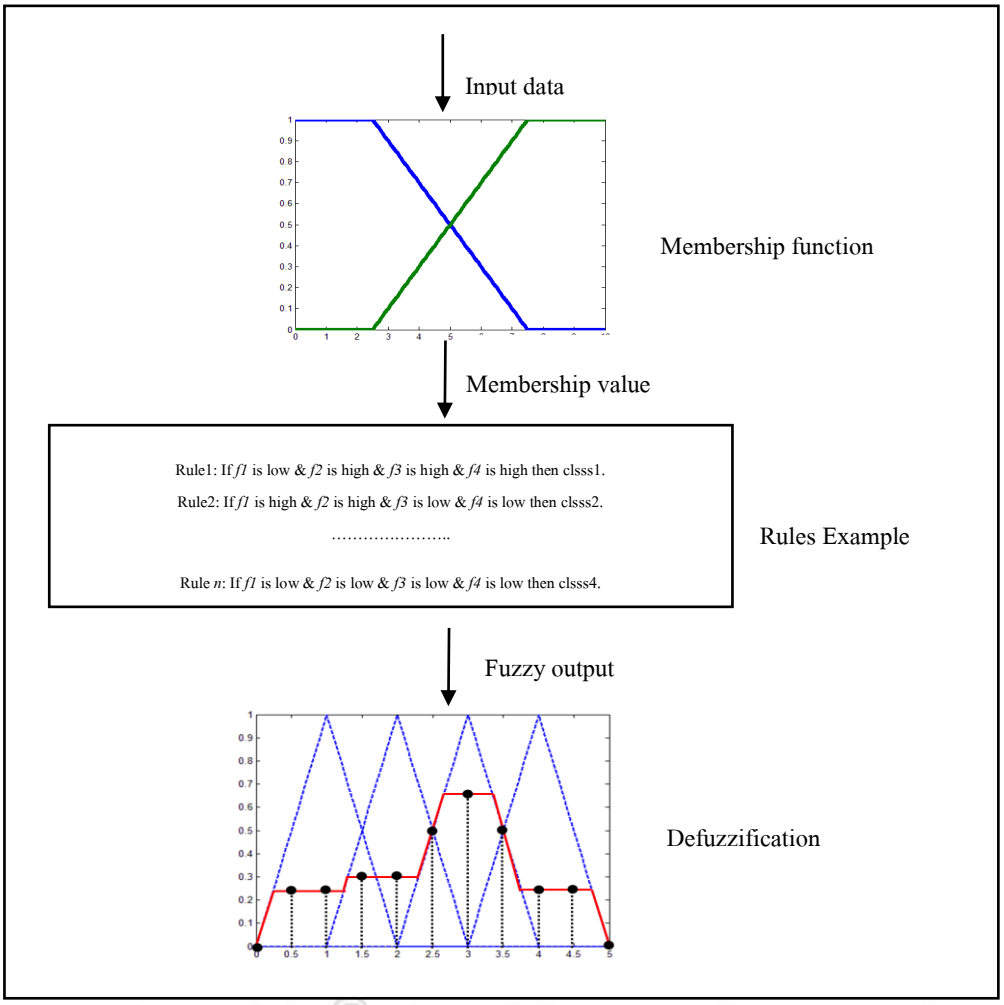
$$M = \left\{ \frac{\sum_{i=1}^{b} y_i}{b} \mid y_i \in \max(\mu(y_i)) \right\} \tag{2.37}$$

เมื่อ $\mu(y_i)$ คือค่าความเป็นสมาชิกของข้อมูลตำแหน่งที่ i

y_i คือค่าจริงของข้อมูลตำแหน่งที่ i

b คือตำแหน่งที่มีดรีกรีความเป็นสมาชิกสูงสุด

ในส่วนของขั้นตอนการทำงานนั้นเป็นไปตามรูปที่ 2.26 โดยจะเริ่มจากการนำข้อมูลมาผ่านฟังก์ชันความเป็นสมาชิก ได้ค่าความเป็นสมาชิกออกมา จากนั้นสร้างกฎขึ้นมาโดยการแปลงข้อมูลจากข้อมูลเชิงตัวเลขเป็นข้อมูลเชิงภาษา ซึ่งจากรูปจะได้ออกมาตามตัวอย่างของกฎ (Rule example) จากนั้นทำการรวมกฎจนสุดท้ายได้ค่าคำตอบออกมา

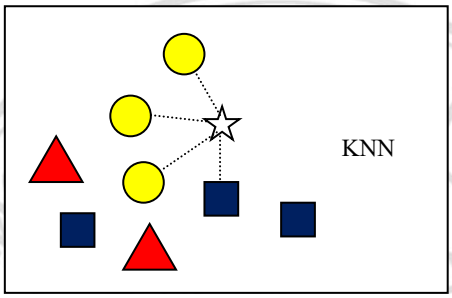


รูปที่ 2.26 ขั้นตอนการทำงานของฟัซซีโลจิก

2.2.2.3 เคเนียเรสเนเบอร์ (K-nearest Neighbor : KNN)[17]

เคเนียเรสเนเบอร์เป็นหนึ่งในอัลกอริทึมที่ถูกใช้กันอย่างกว้างขวางในด้านการจำแนกข้อมูลและการจับกลุ่มข้อมูลโดยใช้ในการระบุคลาส (Class) ของข้อมูล เพราะเป็นอัลกอริทึมที่เข้าใจง่าย มีหลักการทำงานไม่ยุ่งยากซับซ้อน แต่ให้ประสิทธิภาพสูง หลักการทำงานนั้นอาศัยการโหวด (Vote) ว่าข้อมูลที่กำลังพิจารณามีระยะทาง (Distance) ใกล้กับข้อมูลใดๆบ้างจำนวน K ตัว แล้วทำการนับว่าข้อมูล K ตัวนั้นๆอยู่ในคลาสใด เพื่อทำการระบุคลาสของข้อมูลที่กำลังพิจารณา

จากรูปที่ 2.27 แทนรูปดาวด้วยข้อมูลที่กำลังพิจารณาว่าอยู่คลาสใด โดยในรูปจะมีข้อมูลที่รู้คลาสแล้วทั้งสิ้น 8 ข้อมูล แบ่งเป็นรูปสามเหลี่ยม 2 ข้อมูล รูปสี่เหลี่ยม 3 ข้อมูล และ รูปวงกลม 3 ข้อมูล ซึ่งผลจากการใช้ค่า K = 4 พบว่า รูปดาวอยู่ใกล้รูปวงกลม 3 รูปและสี่เหลี่ยม 1 รูป จึงสรุปได้ว่ารูปดาวอยู่ในคลาสวงกลม



รูปที่ 2.27 ตัวอย่างเคเนียเรสเนเบอร์โดยใช้ K = 4

ในการหาระยะทางของข้อมูลนั้นมีหลากหลายวิธีคือ

1) การหาระยะทางโดยพิจารณาจากตัวอักษร (Linguistic Variable Distance)

การหาระยะทางวิธีนี้ต้องใช้สำหรับข้อมูลที่เป็นตัวอักษรเท่านั้น โดยจะทำการนับตัวอักษรที่เหมือนกันว่ามีจำนวนเท่าไหร่ โดยใช้สมการ (2.38)

$$d_{ling}(A,B) = \frac{t-m}{t} \tag{2.38}$$

เมื่อ t คือ จำนวนตัวอักษรทั้งหมด

m คือ จำนวนตัวอีกษรที่เหมือนกัน

$d_{ling}(A,B)$ คือ ระยะทางระหว่างชุดข้อมูล A และ B

2) ระยะทางแบบยุคลิด (Euclidean Distance)

ระยะทางแบบยุคลิดจะเป็นการหาระยะทางตามแนวเส้นตรงของชุดข้อมูล 2 ชุด โดยใช้สมการ (2.39)

$$d_E(A,B) = \sqrt{\sum_{i=1}^{n}(b_i - a_i)^2} \quad (2.39)$$

เมื่อ $d_E(A,B)$ คือระยะทางแบบยุคลิดของชุดข้อมูล A และ B

a_i, b_i คือค่าที่มิติ i ของชุดข้อมูล A และ B ตามลำดับ

3) ระยะทางแบบมินคอฟสกี (Minkowski Distance)

ระยะทางแบบมินคอฟสกีมีลักษณะคล้ายกับระยะทางแบบยุคลิดแต่จะแก้ไขสมการเป็นยกกำลัง และหารากที่ j แทน 2 ดังสมการ (2.40)

$$d_M(A,B) = \sqrt[j]{\sum_{i=1}^{n}(b_i - a_i)^j} \quad (2.40)$$

เมื่อ $d_M(A,B)$ คือระยะทางแบบมินคอฟสกีของ A และ B

a_i, b_i คือค่าที่มิติ i ของชุดข้อมูล A และ B ตามลำดับ

j คือจำนวนเต็มบวก

2.2.2.4 ต้นไม้ตัดสินใจ (Decision Tree) [2]

ต้นไม้ตัดสินใจเป็นเทคนิคในการจำแนกที่ทำความเข้าใจได้ง่าย มีขั้นตอนไม่ยุ่งยากซับซ้อน การสร้างโมเดลของต้นไม้ตัดสินใจนั้นจะทำการหาแอททริบิวต์ที่มีความเกี่ยวข้องกับคลาสมากที่สุดก่อนโดยกำหนดให้เป็นโหนดราก (Root Node) ซึ่งจะอยู่ด้านบนสุดของต้นไม้ จากนั้นหาแอททริบิวต์ที่มีความเกี่ยวข้องกับคลาสลำดับรองๆลงมาโดยใส่ข้อมูลไว้ในแต่ละโหนด ส่วนโหนดล่างสุดของต้นไม้นั้นจะเป็นโหนดสำหรับระบุคลาส ซึ่งรูปแบบการทำงานของต้นไม้ตัดสินใจนั้นจะทำงานจากบนลงล่าง (Top-down)

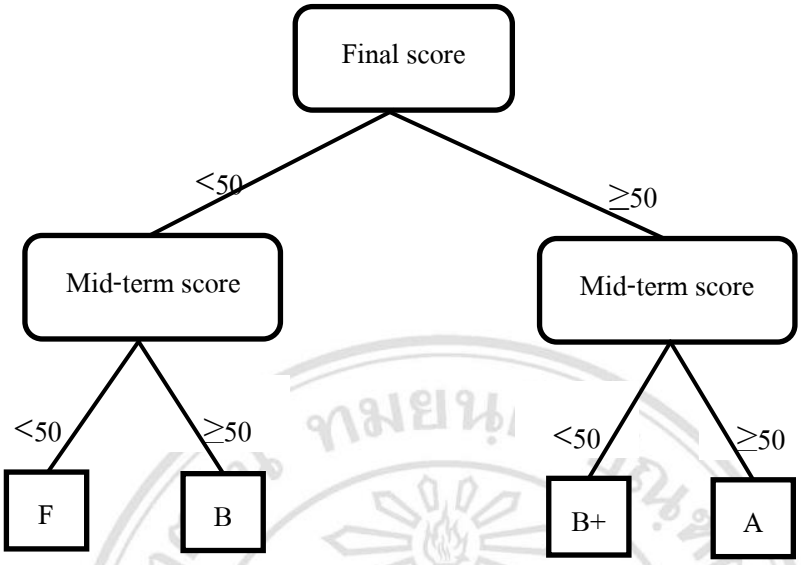
จากรูปที่ 2.28 เป็นตัวอย่างต้นไม้ตัดสินใจโดยใช้หลักการ ถ้า-แล้ว ซึ่งจากรูปสามารถแปลความหมายได้ดังนี้

ถ้า Final score มากกว่าหรือเท่ากับ 50 และ Mid-term score มากกว่าหรือเท่ากับ 50 แล้วจะได้เกรด A

ถ้า Final score มากกว่าหรือเท่ากับ 50 และ Mid-term score น้อยกว่า 50 แล้วจะได้เกรด B+

ถ้า Final score น้อยกว่า 50 และ Mid-term score มากกว่าหรือเท่ากับ 50 แล้วจะได้เกรด B

ถ้า Final score น้อยกว่า 50 และ Mid-term score น้อยกว่า 50 แล้วจะได้เกรด F



รูปที่ 2.28 ตัวอย่างต้นไม้ตัดสินใจ

ซึ่งต้นไม้ตัดสินใจนั้นมีจุดเด่นที่สามารถเข้าใจได้ง่าย แต่ว่าหากมีข้อมูลจำนวนมากขะทำให้การสร้างต้นไม้ตัดสินใจเป็นไปได้ลำบาก

2.2.2.5 นาอีฟเบย์ (Naïve Bayes) [2]

นาอีฟเบย์เป็นหนึ่งในเทคนิคที่ได้รับความนิยมในการทำเหมืองข้อมูล เพราะเป็นเทคนิคที่ง่ายต่อการเข้าใจ ไม่ยุ่งยากซับซ้อน มีความรวดเร็วในการทำงาน และสามารถเพิ่มความสามารถในการทำงานได้โดยการเพิ่มจำนวนข้อมูลที่ใช้ในการสอนเข้าไป นาอีฟเบย์นั้นอาศัยหลักการเกี่ยวกับความน่าจะเป็น (Probability) ในการทำงานโดยอาศัยสมการ (2.41) ซึ่งสมการนี้ถูกเรียกว่า ทฤษฎีของเบย์ (Bayes Theorem)

$$P(C \mid A) = \frac{P(A \mid C) * P(C)}{P(A)} \qquad (2.41)$$

เมื่อ $P(C \mid A)$ คือค่าความน่าจะเป็นที่ค่าของข้อมูลแอททริบิวต์เป็น A จะมีคลาส C

$P(A \mid C)$ คือค่าความน่าจะเป็นที่ค่าของคลาสเป็น C จะมีค่าของข้อมูลแอททริบิวต์เป็น A

$P(C)$ คือค่าความน่าจะเป็นที่ค่าของคลาสเป็น C

$P(A)$ ค่าความน่าจะเป็นที่ค่าของข้อมูลแอททริบิวต์เป็น A

แต่เทคนิคนาอีฟเบย์นั้นเมื่อข้อมูลมีความซับซ้อนมากขึ้น จำเป็นต้องมีข้อมูลที่หลากหลายในการสร้างโมเดล ทำให้ไม่ค่อยเหมาะสมกับการทำงานที่ต้องการความซับซ้อน แต่ในงานที่ไม่มีความซ้ำซ้อนมาก นาอีฟเบย์จำเป็นตจะให้ประสิทธิภาพที่ดีมาก

2.2.2.6 เกาส์เซียนมิกซ์เจอร์โมเดล (Gaussian Mixture Model: GMM)[18]

เกาส์เซียนมิกซ์เจอร์โมเดลเป็นอีกหนึ่งเทคนิคที่สามารถนำมาใช้ในการจำแนกข้อมูลได้ เกาส์เซียนมิกซ์เจอร์โมเดลนั้นอาศัยหลักการการกระจายตัวแบบเกาส์เซียน (Gaussian Distribution) ในการจำแนกข้อมูลที่มีหลายแอททริบิวต์นั้นจำเป็นต้องรวมกราฟเกาส์เซียนหลายๆกราฟเข้าด้วยกันเพื่อให้ได้กราฟที่เหมาะสมก่อนนำไปใช้ในการจำแนก ซึ่งสมการ (2.42), (2.43) และ (2.44) ใช้ในการสร้างเกาส์เซียนมิกซ์เจอร์โมเดลแบบหลายแอททริบิวต์ของแต่ละคลาสเพื่อใช้ในการจำแนก

$$\alpha \approx \frac{1}{N}\sum_{i=1}^{N} x_i \tag{2.42}$$

เมื่อ N คือจำนวนข้อมูลทั้งหมดในคลาสนั้นๆ

x คือข้อมูลทั้งหมดในแต่ละคลาสเพื่อใช้สร้างโมเดล

$$\beta(p,q) \approx \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i(p)-a(p)\right)*\left(x_i(q)-a(q)\right) \tag{2.43}$$

เมื่อ p,q คือตำแหน่งของแถวและหลัก

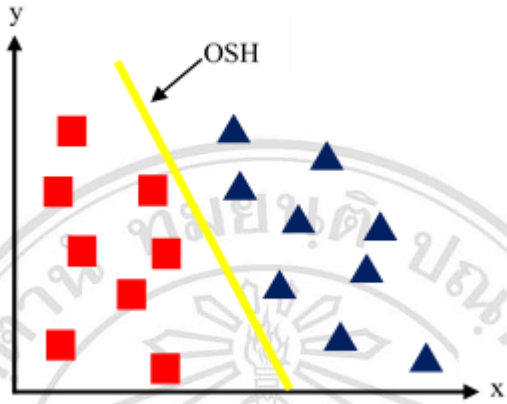
$$f_G(x \mid \alpha,\beta) = \frac{1}{\sqrt{(2\pi)^k \det(\beta)}} e^{-\frac{1}{2}(x-\alpha)^T \beta^{-1}(x-\alpha)} \tag{2.44}$$

เมื่อ f_G คือโอกาสในการเกิดคลาส

2.2.2.7 ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) [19]

ซัพพอร์ตเวกเตอร์แมชชีนเป็นเทคนิคในด้านการจำแนกข้อมูลเชิงเส้น (Linear Classifier) ที่ได้รับความนิยมค่อนข้างมาก เพราะมีประสิทธิภาพในการทำงานที่สูง และมักจะไม่พบเจอปัญหาโอเวอร์ฟิตติ่ง (Overfitting) ซึ่งโอเวอร์ฟิตติ่งนั้นคือเหตุการณ์ที่โมเดลที่สร้างขึ้นมาได้รับการเรียนรู้มากเกินไปจนไม่สามารถนำโมเดลที่สร้างมาจำแนกข้อมูลใหม่ได้

ซัพพอร์ตเวกเตอร์แมชชีนนั้นอาศัยหลักการการหาระนาบแบ่งแยกที่ดีที่สุด (Optimal Separation Hyperplane: OSH) โดยระราบแบ่งแยกที่ดีที่สุดนี้จะใช้เป็นเส้นแบ่งของคลาสข้อมูล โดย รูปที่ 2.29 แสดงตัวอย่างการใช้ระนาบแบ่งแยกแบ่งข้อมูลสองคลาส



รูปที่ 2.29 ตัวอย่างการใช้งานซัพพอร์ตเวกเตอร์แมชชีน

แต่ในความเป็นจริงนั้นข้อมูลที่นำมาใช้งานมักจะไม่ใช่ข้อมูลเชิงเส้นจึงต้องมีการนำเคอร์เนลฟังก์ชัน (Kernel Function) มาช่วยในการทำงาน

2.3 ภาษาทารกของดันส์ตัน (Dunstan Baby Language) [1]

เสียงร้องของเด็กทารกตั้งแต่แรกเกิดถึงสามเดือนนั้นเป็นเสียงร้องช่วงก่อนที่เด็กจะได้เรียนรู้ภาษา จึงเป็นปฏิกิริยาตอบสนองที่เกิดขึ้นเองของร่างกาย ดังนั้นเด็กทารกทุกคนจึงมีเสียงร้องเหล่านี้คล้ายคลึงกัน ไม่ว่าจะเป็นชนชาติอะไร เสียงเหล่านี้แบ่งออกเป็น 5 คำคือ

1. เนะ (Neh) เป็นเสียงที่ทารกรู้สึกหิว เป็นเสียงที่เกิดจากลิ้นไปผลักด้านบนของเพดานปาก เมื่อได้ยินเสียงนี้คนที่ดูแลเด็กอยู่ควรหาอาหารให้ทารกรับประทาน

2. อาว (Owh) เป็นเสียงที่ทารกรู้สึกเหนื่อย อยากนอนพัก เสียงนี้มีลักษณะคล้ายการหาว เมื่อได้ยินเสียงนี้คนที่ดูแลเด็กอยู่ควรพาเด็กเข้านอน

3. เฮะ (Heh) เป็นเสียงที่ทารกรู้สึกไม่สบายตัว หรืออยากเปลี่ยนผ้าอ้อมใหม่ เสียงนี้เกิดจากการตอบสนองบริเวณผิวหนังเช่น เหงื่อ เมื่อได้ยินเสียงนี้คนที่ดูแลเด็กอยู่ควรเปลี่ยนผ้าอ้อมอันใหม่ หรือเช็ดเหงื่อตามตัวเด็ก

4. แอะ (Eairh) เป็นเสียงที่ทารกรู้สึกปวดท้องหรือท้องอืด เสียงนี้เกิดจากมีอากาศขังอยู่จากการเรอ โดยอากาศไม่สามารถเดินทางไปยังกระเพาะอาหาร เพื่อให้ลำไส้ปล่อยฟองอากาศออกไป เมื่อได้ยินเสียงนี้คนที่ดูแลเด็กอยู่ควรทำให้เด็กเรอออกมา

5. เอะ (Eh) เป็นเสียงที่ทารกรู้สึกอยากเรอ เสียงนี้เกิดจากมีฟองอากาศตกค้างอยู่ในหน้าอกและต้องการปล่อยออกมาทางช่องปาก เมื่อได้ยินเสียงนี้คนที่ดูแลเด็กอยู่ควรทำให้เด็กเรอออกมา

2.4 การตรวจสอบความสมเหตุสมผลแบบไขว้ (Cross-Validation)

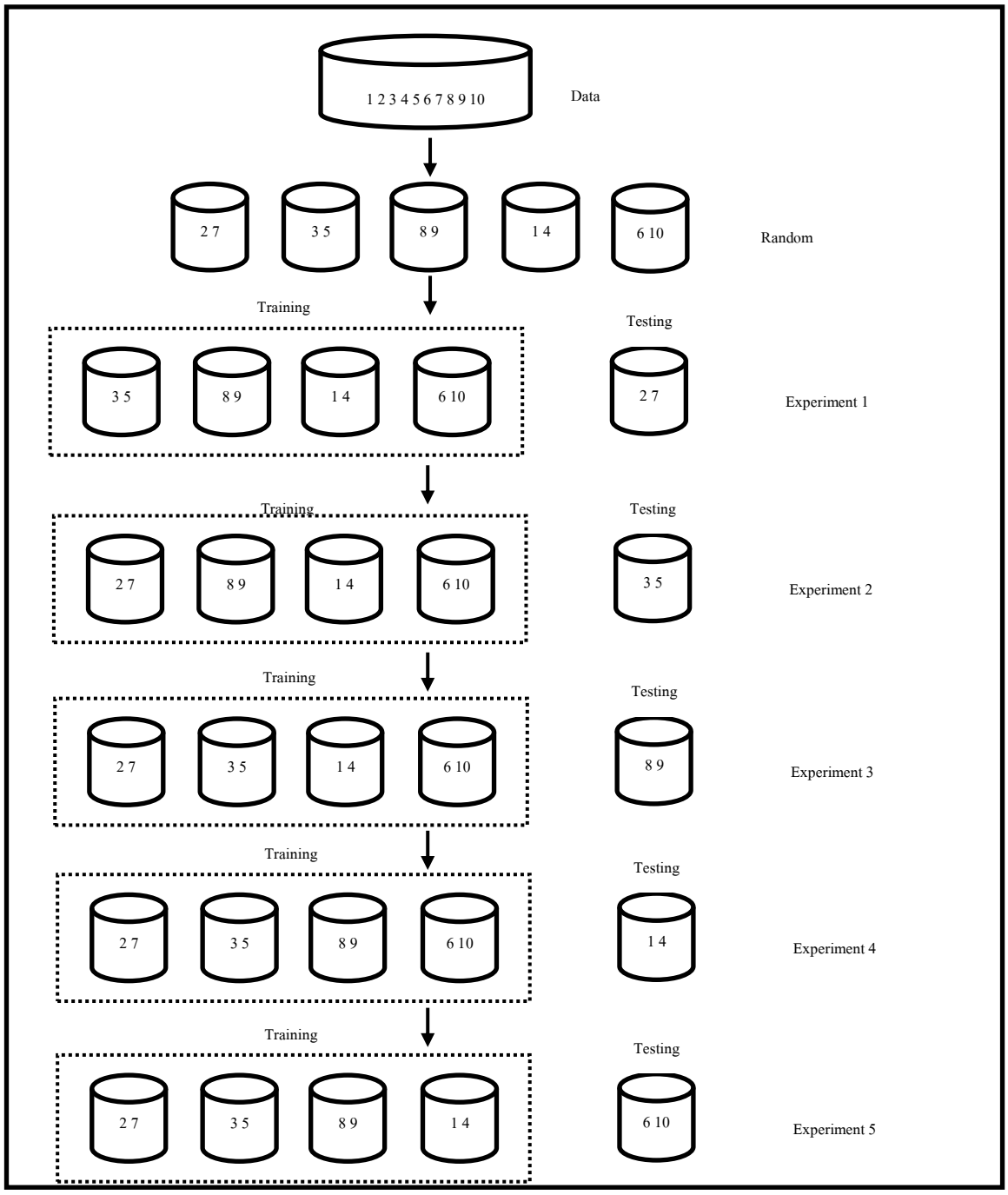
การตรวจสอบความสมเหตุสมผลแบบไขว้เป็นกระบวนการทางสถิติเพื่อใช้ในการวิเคราะห์ผลของคำตอบที่ได้ ซึ่งเป็นวิธีที่ได้รับการยอมรับเพราะข้อมูลทุกตัวจะได้เป็นทั้งข้อมูลที่ใช้เรียนรู้และข้อมูลที่ใช้ทดสอบ โดยผลลัพธ์ที่ได้จากการบวนการตรวจสอบความสมเหตุสมผลแบบไขว้จะเป็นการหาค่าเฉลี่ยรวมของทุกๆการทดลอง

ในการทำงานนั้นจะแบ่งข้อมูลออกเป็น k กลุ่มแบบสุ่ม เรียกว่า k-fold Cross-Validation โดยจะทำการทดลองทั้งหมด k ครั้ง ในแต่ละครั้งจะใช้ข้อมูลเรียนรู้ k-1 กลุ่ม และใช้ข้อมูลทดสอบ 1 กลุ่ม

จากรูปที่ 2.30 จะเริ่มจากทำการสุ่มข้อมูลเข้ากลุ่มที่ 1-5 จากนั้นแบ่งการทดลองออกเป็น 5 ครั้งคือ

- ครั้งที่ 1 ใช้ข้อมูลชุดที่ 2 ถึง 4 ในการสอนระบบ และใช้ข้อมูลชุดที่ 1 ทดสอบระบบ
- ครั้งที่ 2 ใช้ข้อมูลชุดที่ 1 3 4 และ 5 ในการสอนระบบ และใช้ข้อมูลชุดที่ 2 ทดสอบระบบ
- ครั้งที่ 3 ใช้ข้อมูลชุดที่ 1 2 4 และ 5 ในการสอนระบบ และใช้ข้อมูลชุดที่ 3 ทดสอบระบบ
- ครั้งที่ 4 ใช้ข้อมูลชุดที่ 1 ถึง 3 และ 5 ในการสอนระบบ และใช้ข้อมูลชุดที่ 4 ทดสอบระบบ
- ครั้งที่ 5 ใช้ข้อมูลชุดที่ 1 ถึง 4 ในการสอนระบบ และใช้ข้อมูลชุดที่ 5 ทดสอบระบบ

สุดท้ายแล้วผลจาก 5-fold Cross-Validation จะได้ผลจากการทดลองที่ 1 ถึง 5 มาหาค่าเฉลี่ยออกมาเป็นคำตอบ



รูปที่ 2.30 แสดงการแบ่งข้อมูลแบบ 5-fold Cross-Validation