บทที่ 9

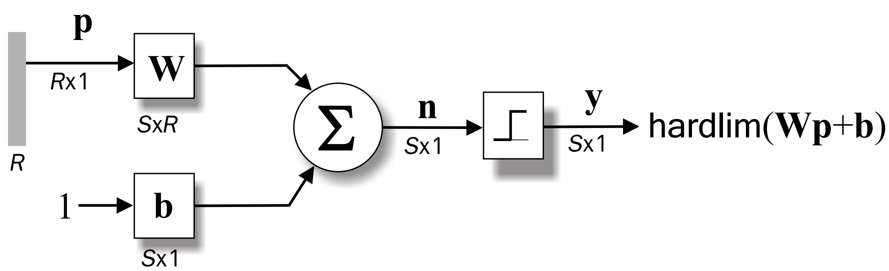
การเรียนรู้ของเครือข่ายเพอร์เซ็พตรอน
Perceptron Learning

Warren McCulloch และ Walter Pitts [McCulloch and Pitts, 1943] ได้นำเสนอแบบจำลองนิวรอนรุ่นแรกในปี 1943 โดยมีหลักการง่ายในการรวมอินพุตพร้อมทั้งค่าน้ำหนักประสาท แล้วนำไปเปรียบเทียบกับค่าจุดเปลี่ยนระดับ(thresholding) ซึ่งถ้าผลรวมมีค่ามากกว่าจุดเปลี่ยนระดับ เอาต์พุตของนิวรอนจะให้ค่าเป็น 1 ถ้าผลรวมมีค่าไม่ถึงจุดเปลี่ยนระดับ เอาต์พุตของนิวรอนจะมีค่า 0 ในการนำเสนอได้มีการแสดงถึงความสามารถในการคำนวณทางคณิตศาสตร์ได้ ในการนำเสนอไม่ได้กล่าวถึงการเรียนรู้ของเครือข่าย ค่าพารามิเตอร์ต่างๆ ถูกกำหนดขึ้นเอง อย่างไรก็ตามโครงสร้างที่ได้นี้เป็นการเชื่อมโยงระหว่างระบบประสาทของมนุษย์ กับแบบจำลองประสาทที่ได้รับการพัฒนาต่อมาจนกระทั่งถึงปัจจุบัน

ต่อมาในช่วงปลายศตวรรษที่ 1950 Frank Rosenblatt [Rosenblatt, 1960, 1962, 1958] และนักวิจัยค้นคว้าอีกมากมายได้พัฒนาเครือข่ายประสาทเทียมเรียกว่าเพอร์เซ็พตรอน(perceptron) ซึ่งมีโครงสร้างที่คล้ายคลึงกับแบบจำลองของ McCulloch และ Pitts โดยได้นำเสนอถึงวิธีในการเรียนรู้ของเครือข่าย ในการนำไปประยุกต์แก้ปัญหางานทางด้านการจดจำรูปแบบ (pattern recognition) Rosenblatt ได้พิสูจน์ว่ากฎการเรียนรู้ที่ได้นำเสนอนั้นจะลู่เข้าสู่ค่าน้ำหนักประสาทของเครือข่ายได้อย่างถูกต้องเสมอ (นั่นคือสามารถฝึกสอนเครือข่ายจนกระทั่งเรียนรู้และใช้งานได้) การเรียนรู้ของเครือข่ายไม่มีความซับซ้อนและเป็นไปอย่างอัตโนมัติ นอกจากนั้นยังได้แสดงถึงตัวอย่างพฤติกรรมของระบบของเครือข่าย ที่สามารถเรียนรู้จากข้อผิดพลาดของตัวเองได้ เพอร์เซ็พตรอนที่ถูกนำเสนอยังมีความสามารถในการเรียนรู้ ถึงแม้ว่าค่าน้ำหนักประสาทและไบอัสในสภาวะเริ่มต้นจะเป็นค่าจากการสุ่ม

อย่างไรก็ตาม ตัวเพอร์เซ็พตรอนนั้นก็มีข้อจำกัดอยู่หลายๆ ด้าน ซึ่งในเวลาต่อมาได้มีการพัฒนาโครงสร้างรูปแบบของเพอร์เซ็พตรอน เพื่อแก้ไขข้อจำกัดต่างๆ เหล่านั้น เครือข่ายเพอร์เซ็พตรอนยังคงมีความสำคัญและถูกนำมาใช้งานจนกระทั่งถึง ณ ปัจจุบัน ซึ่งถือเป็นเครือข่ายที่มีความเร็วสูงและมีความน่าเชื่อถือ (reliable) ในระดับของงานที่พิสูจน์แล้วว่าเพอร์เซ็พตรอนสามารถรองรับได้ เนื่องจากเครือข่ายเพอร์เซ็พตรอนไม่มีโครงสร้างที่ซับซ้อน รวมไปถึงกฎการเรียนรู้ที่ง่ายและตรงไปตรงมา จึงเป็นจุดเริ่มต้นที่ดีในการศึกษาในด้านพื้นฐานเกี่ยวกับเครือข่ายประสาทเทียมต่อไป

เนื้อหาในบทนี้จะได้กล่าวถึงสถาปัตยกรรมของเพอร์เซ็พตรอน และกฎการเรียนรู้แบบเพอร์เซ็พตรอน รวมไปถึงตัวอย่างการปรับแต่งกฎในรูปแบบต่างๆ เพื่อให้ได้การเรียนรู้ที่มีประสิทธิภาพมากขึ้น



รูปที่ 9.1: เครือข่ายเพอร์เซ็พตรอน

9.1 สถาปัตยกรรมของเพอร์เซ็พตรอน

โครงสร้างทั่วๆ ไปของเพอร์เซ็พตรอนมีแสดงในรูปที่ 9.1 เอาต์พุตของเครือข่ายคือ

$$\mathbf{y} = \text{hardlim}(\mathbf{W}\mathbf{p} + \mathbf{b}) \tag{9.1}$$

เมื่อพิจารณารายละเอียดของเครือข่ายเพอร์เซ็พตรอนดังกล่าว จะเห็นว่าเครือข่ายเพอร์เซ็พตรอนคือเครือข่ายแบบไปข้างหน้า (feedforward network) ที่มีจำนวนชั้นของนิวรอนเพียง 1 ชั้น มีจำนวนนิวรอนเท่ากับ S และฟังก์ชันถ่ายโอนเป็นแบบฮาร์ดลิมิต (เอาต์พุตมีค่าเป็น 0 หรือ 1) อินพุตของเครือข่ายมีจำนวนเท่ากับ R และมีจำนวนไบอัสของแต่ละนิวรอนรวมแล้วเท่ากับ S ในเนื้อหาถัดไป ผู้อ่านจะได้วิเคราะห์ในเรื่องข้อจำกัดของเครือข่ายที่มีสถาปัตยกรรมของนิวรอนเพียง 1 ชั้นนี้

พิจารณาเมตริกซ์น้ำหนักประสาท

$$\mathbf{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1R} \\ w_{21} & w_{22} & \cdots & w_{2R} \\ \vdots & \vdots & & \vdots \\ w_{S1} & w_{S2} & \cdots & w_{SR} \end{bmatrix} \tag{9.2}$$

เพื่อความสะดวกในบางครั้ง เราจะเขียนแยกให้อยู่ในรูปเมตริกซ์แถวที่ i ของ $\mathbf{W}$ คือ

$$_i\mathbf{w} = \begin{bmatrix} w_{i1} & w_{i2} & \cdots & w_{iR} \end{bmatrix} \tag{9.3}$$

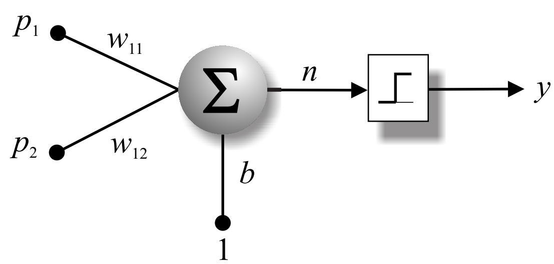
ดังนั้นสามารถเขียนเมตริกซ์น้ำหนักประสาท $\mathbf{W}$ ในรูปของเมตริกซ์แถว $_i\mathbf{w}$ ได้ดังนี้

$$\mathbf{W} = \begin{bmatrix} _1\mathbf{w} \\ _2\mathbf{w} \\ \vdots \\ _S\mathbf{w} \end{bmatrix} \tag{9.4}$$

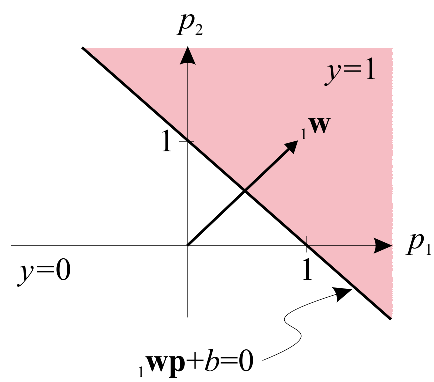
เราสามารถเขียนเอาต์พุตของเครือข่ายแยกเป็นองค์ประกอบย่อยตามเมตริกซ์แถวที่ i ได้คือ

$$y_i = \text{hardlim}(n_i) = \text{hardlim}(_i\mathbf{w}\mathbf{p} + b_i) \tag{9.5}$$

จะได้ว่าถ้าผลคูณภายใน (inner product) ของเมตริกซ์น้ำหนักประสาทแถวที่ i กับเวกเตอร์ของอินพุตมากกว่าหรือเท่ากับ $-b_i$ แล้ว เอาต์พุตของเครือข่ายจะมีค่าเป็น 1 ถ้าค่าผลคูณภายในเป็นนอกเหนือจากนี้แล้ว เอาต์พุตของเครือข่ายจะเป็น 0 ผลสรุปที่สำคัญ ณ จุดนี้คือ *นิวรอนแต่ละตัวในเครือข่ายจะทำการแบ่งพื้นที่ของอินพุต (input space) ออกเป็นสองพื้นที่* ในหัวข้อต่อไปจะทำการศึกษาเกี่ยวกับเส้นแบ่งพื้นที่ดังกล่าว โดยจะพิจารณาในกรณีของเครือข่ายที่มีนิวรอนตัวเดียวและมี 2 อินพุต



รูปที่ 9.2: เพอร์เซ็พตรอนนิวรอนเดียวแบบ 2 อินพุต



รูปที่ 9.3: ตัวอย่างเส้นแบ่งพื้นที่ในกรณี $_1\mathbf{w} = [1 \quad 1]$

9.1.1 เพอร์เซ็พตรอนแบบนิวรอนเดียว Single-Neuron Perceptron

พิจารณาโครงสร้างของเพอร์เซ็พตรอนที่มีเพียงนิวรอนเดียวในรูปที่ 9.2 เอาต์พุตของเครือข่ายสามารถเขียนได้ดังนี้ (พิจารณาใช้สัญลักษณ์ $_1\mathbf{w}$ แทนเวกเตอร์น้ำหนักประสาทแถวที่ 1 ของเครือข่ายเพอร์เซ็พตรอนที่มี 1 นิวรอน)

$$\begin{aligned} y &= \text{hardlim}(n) = \text{hardlim}(\mathbf{Wp}+b) \\ &= \text{hardlim}(_1\mathbf{wp}+b) \\ &= \text{hardlim}(w_{11}p_1 + w_{12}p_2 + b) \end{aligned} \quad (9.6)$$

ความสัมพันธ์ข้างต้นสามารถพิจารณาเป็นการแบ่งพื้นที่ของอินพุต (p_1, p_2) ออกเป็น 2 พื้นที่ได้ด้วย**เส้นแบ่งพื้นที่** (decision boundary) โดยพิจารณาจากความสัมพันธ์ของเอาต์พุต n มีค่าเท่ากับศูนย์ดังนี้

$$n =_1 \mathbf{wp} + b = w_{11}p_1 + w_{12}p_2 + b = 0 \quad (9.7)$$

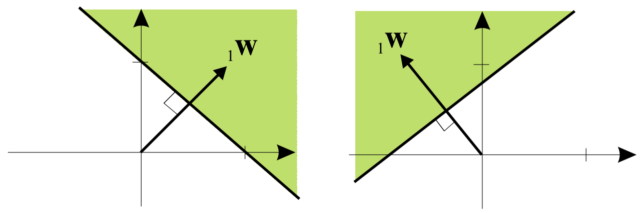
พิจารณาแทนค่าน้ำหนักประสาทและไบอัสด้วย $w_{11} = 1$ $w_{12} = 1$ และ $b = -1$ จะได้เส้นแบ่งพื้นที่คือ

$$n =_1 \mathbf{wp} + b = w_{11}p_1 + w_{12}p_2 + b = p_1 + p_2 - 1 = 0 \quad (9.8)$$

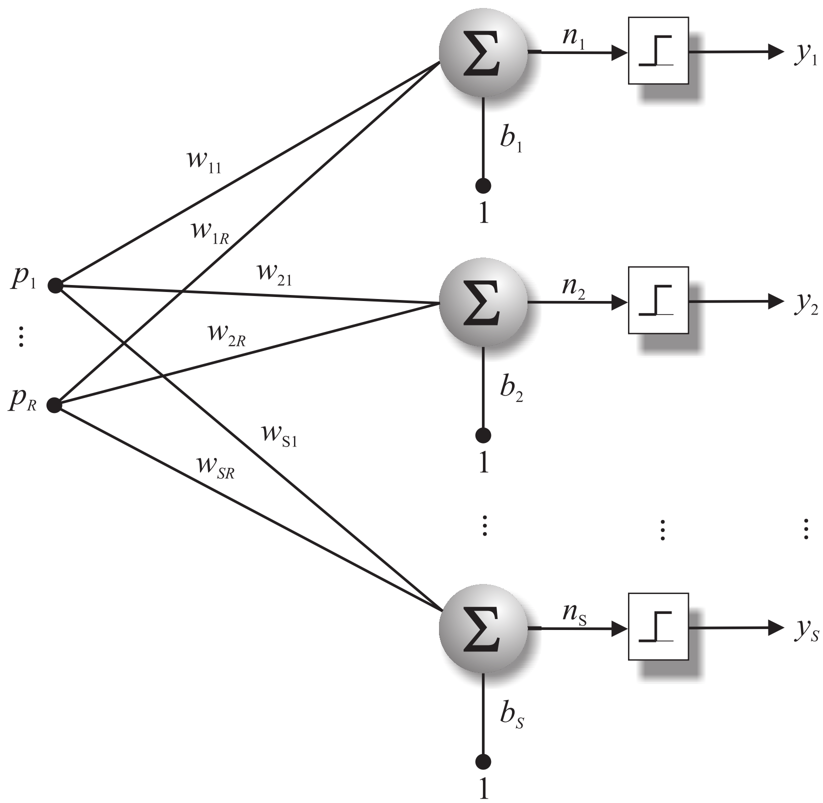
รูปที่ 9.3 แสดงเส้นแบ่งพื้นที่ โดยพื้นที่ตั้งแต่เส้นตรงและส่วนแรเงาขึ้นไปเป็นส่วนที่ $n \geq 0$ ซึ่งให้ค่าเอาต์พุตของเครือข่ายเป็น $y = 1$ สังเกตว่าเส้นแบ่งพื้นที่จะตั้งฉากกับ $_1\mathbf{w}$ เสมอดังตัวอย่างแสดงในรูปที่ 9.4 $_1\mathbf{w}$ จะชี้ไปในทิศทางของพื้นที่ที่ซึ่งเอาต์พุต y ของนิวรอนมีค่าเป็น 1 เส้นแบ่งพื้นที่สามารถนิยามได้ดังนี้

$$_1\mathbf{wp} + b = 0 \quad (9.9)$$

นิวรอนที่ใช้ในการแบ่งพื้นที่ข้างต้น สามารถแบ่งพื้นที่ออกเป็นสองส่วน ในกรณีของเพอร์เซ็พตรอนที่มีหลายนิวรอน เส้นแบ่งพื้นที่ดังกล่าวจะมีจำนวนเท่ากับจำนวนนิวรอนด้วย ความสามารถในการแบ่งพื้นที่ปริภูมิอินพุตได้



รูปที่ 9.4: เส้นแบ่งพื้นที่ตั้งฉากกับเมตริกซ์น้ำหนักประสาทเสมอ



รูปที่ 9.5: เพอร์เซ็พตรอนแบบหลายนิวรอน

หลายๆ พื้นที่มีความจำเป็นต่อปัญหาที่มีความซับซ้อนมากขึ้นในหลายๆ กรณี โดยเฉพาะงานทางด้านการจำแนกหรือจดจำรูปแบบ นิวรอนที่กล่าวถึงในหัวข้อข้างต้น เป็นนิวรอนที่มี 2 อินพุต การแสดงการแบ่งพื้นที่สามารถแสดงได้ด้วยรูป 2 มิติ ในทางปฏิบัติแล้ว จำนวนอินพุตที่ใช้งานจริงจะมากกว่า 2 การแสดงผลของตัวแปรมากกว่า 2 นั้นไม่สะดวก เนื้อหาในบทนี้จะใช้นิวรอนขนาด 2 อินพุตในการแสดงผล รายละเอียดของเพอร์เซ็พตรอนแบบหลายนิวรอนจะได้กล่าวถึงในหัวข้อต่อไปนี้

9.1.2 เพอร์เซ็พตรอนแบบหลายนิวรอน Multiple-Neuron Perceptron

สำหรับเพอร์เซ็พตรอนแบบหลายนิวรอนดังแสดงในรูปที่ 9.5 แต่ละนิวรอนจะมีเส้นแบ่งพื้นที่ โดยเส้นแบ่งพื้นที่ของนิวรอนตัวที่ i คือ

$$ {}_i\mathbf{w}\mathbf{p} + b_i = 0 \tag{9.10} $$

เพอร์เซ็พตรอนแบบนิวรอนเดียวสามารถคัดแยกอินพุตเวกเตอร์ออกเป็นสองกลุ่ม ที่มีเอาต์พุตค่าเป็น 0 และ 1 ดังนั้นเพอร์เซ็พตรอนแบบหลายนิวรอนจะสามารถคัดแยกอินพุตออกเป็นหลายๆ กลุ่มได้ แต่ละกลุ่มมีเอาต์พุตเวกเตอร์ที่ต่างกัน จำนวนกลุ่มที่เพอร์เซ็พตรอน S นิวรอนสามารถคัดแยกได้คือ 2^S (แต่ละนิวรอนสามารถแบ่งพื้นที่ได้ 2 ส่วน)

9.2 กฎการเรียนรู้แบบเพอร์เซ็พตรอน
Perceptron Learning Rule

ในการออกแบบเครือข่ายเพอร์เซ็พตรอน สิ่งที่ต้องทำการคำนวณหาคือค่าของเมตริกซ์น้ำหนักประสาทโดยใช้กฎการเรียนรู้แบบเพอร์เซ็พตรอน ซึ่งเป็นการเรียนรู้แบบมีผู้ฝึกสอน(supervised learning) โดยการนำเสนอคู่อินพุตและเป้าหมายที่ต้องการให้กับเครือข่าย

$$\{\mathbf{p}_1, \mathbf{t}_1\}\{\mathbf{p}_2, \mathbf{t}_2\} \ldots \{\mathbf{p}_Q, \mathbf{t}_Q\} \tag{9.11}$$

โดยที่ $\mathbf{p}_i$ เป็นอินพุตตัวที่ i และ $\mathbf{t}_i$ เป็นคู่เป้าหมายของอินพุต i นั้นๆ อินพุต $\mathbf{p}_i$ แต่ละตัวจะถูกป้อนให้กับเครือข่าย แล้วเอาต์พุต $\mathbf{y}$ ที่ได้จะถูกนำไปเปรียบเทียบกับเป้าหมาย $\mathbf{t}_i$ เครือข่ายจะทำการปรับค่าน้ำหนักประสาทและไบอัสตามกฎการเรียนรู้ เพื่อทำให้เอาต์พุตของเครือข่ายถูกปรับให้เข้าใกล้เป้าหมายมากที่สุด กฎการเรียนรู้แบบเพอร์เซ็พตรอนมีดังนี้ (สัญลักษณ์ ${}_1\mathbf{w}$ คือเวกเตอร์แถวที่ 1 ของเมตริกซ์น้ำหนักประสาท ซึ่งใช้แทนเครือข่ายเพอร์เซ็พตรอนที่มี 1 นิวรอน)

▷ **กฎการเรียนรู้ของเพอร์เซ็พตรอน – I**

$$\begin{array}{llll} \text{ถ้า} & t=1 \text{ และ } y=0 & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} + \mathbf{p} \\ \text{ถ้า} & t=0 \text{ และ } y=1 & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} - \mathbf{p} \\ \text{ถ้า} & t=y & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} \end{array} \tag{9.12}$$

กฎการเรียนรู้ข้างต้นสามารถเขียนให้อยู่ในรูปทั่วไปได้ โดยกำหนดให้ค่าความผิดพลาด (error) ของเครือข่ายเพอร์เซ็พตรอนคือ $e = t - y$ จะได้ว่า

▷ **กฎการเรียนรู้ของเพอร์เซ็พตรอน – II**

$$\begin{array}{llll} \text{ถ้า} & e=1 & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} + \mathbf{p} \\ \text{ถ้า} & e=-1 & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} - \mathbf{p} \\ \text{ถ้า} & e=0 & \text{แล้ว} & {}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} \end{array} \tag{9.13}$$

หรือเขียนให้อยู่ในรูปความสัมพันธ์เดียวได้คือ

$${}_1\mathbf{w}^{new} = {}_1\mathbf{w}^{old} + e\mathbf{p} \tag{9.14}$$

$$= {}_1\mathbf{w}^{old} + (t-y)\mathbf{p} \tag{9.15}$$

สำหรับไบอัสจะใช้กฎการเรียนรู้เดียวกันกล่าวคือ

$$b^{new} = b^{old} + e \tag{9.16}$$

เช่นเดียวกันกับเครือข่ายเพอร์เซ็พตรอนแบบนิวรอนเดียว กฎการเรียนรู้แบบเพอร์เซ็พตรอนในการปรับค่าเมตริกซ์น้ำหนักประสาทแถวที่ i (นั่นคือน้ำหนักประสาทของนิวรอนตัวที่ i) คือ

$$ {}_i\mathbf{w}^{new} =_i \mathbf{w}^{old} + e_i\mathbf{p} \tag{9.17}$$

และสำหรับไบอัสคือ

$$ b_i^{new} = b_i^{old} + e_i \tag{9.18}$$

กฎการเรียนรู้แบบเพอร์เซ็พตรอนสำหรับเครือข่ายหลายนิวรอนสามารถเขียนรวมในรูปของเมตริกซ์น้ำหนักประสาทได้คือ

$$ \mathbf{W}^{new} = \mathbf{W}^{old} + \mathbf{ep} \tag{9.19}$$

และสำหรับไบอัสคือ

$$ \mathbf{b}^{new} = \mathbf{b}^{old} + \mathbf{e} \tag{9.20}$$

ตัวอย่างต่อไปนี้แสดงการทำงานและการเรียนรู้ของเครือข่ายเพอร์เซ็พตรอน โดยเลือกใช้ตัวอย่างการเรียนรู้ของเพอร์เซ็พตรอนเพื่อหาฟังก์ชันแทนตัวปฏิบัติการ OR ฟังก์ชันดังกล่าวถือว่าเป็นฟังก์ชันแบบไม่เป็นเชิงเส้น และไม่มีผลเฉลยรูปแบบปิดใดๆ ที่สามารถใช้แทนฟังก์ชันดังกล่าวได้ ตัวอย่างนี้แสดงให้เห็นถึงความสามารถของเครือข่ายประสาทเทียม ในอีกมุมมองหนึ่งที่ได้รับความสนใจเป็นอย่างมาก ผู้อ่านเองควรทำความเข้าใจตัวอย่างดังกล่าวให้ถ่องแท้ เพื่อให้คุ้นเคยกับลักษณะการทำงานของเครือข่ายประสาทเทียมต่อไป เนื่องจากการทำงานของเครือข่ายประสาทเทียม จะมีความแตกต่างไปจากวิธีการคำนวณแบบดั้งเดิมค่อนข้างมาก

■ **ตัวอย่างที่ 9.1** การเรียนรู้ของเครือข่ายเพอร์เซ็พตรอน

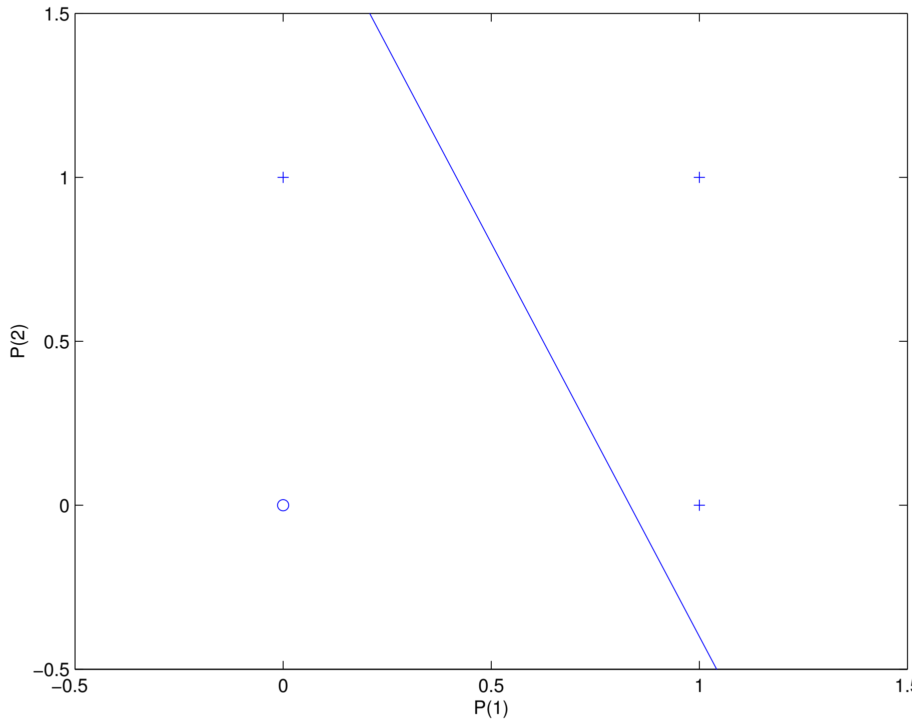
พิจารณาปัญหาการฝึกสอนเพื่อแยกแยะของตัวปฏิบัติการ OR โดยมีเวกเตอร์อินพุตและเวกเตอร์เป้าหมายดังนี้ (ดูรูปที่ 9.6)

$$\left\{\mathbf{p_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{t_1} = [0]\right\}$$
$$\left\{\mathbf{p_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{t_2} = [1]\right\}$$
$$\left\{\mathbf{p_3} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{t_3} = [1]\right\}$$
$$\left\{\mathbf{p_4} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{t_4} = [1]\right\}$$

โดยที่ $\mathbf{p_1} - \mathbf{p_4}$ คือเวกเตอร์อินพุตและ $\mathbf{t_1} - \mathbf{t_4}$ เป็นเวกเตอร์เป้าหมายของ $\mathbf{p_1} - \mathbf{p_4}$ ตามลำดับ ในรูปยังแสดงเส้นแบ่งพื้นที่จากเวกเตอร์น้ำหนักประสาท ${}_1\mathbf{w}$ ซึ่งมีการกำหนดค่าเริ่มต้นคือ ${}_1\mathbf{w} = [-1.2, -0.5]^T$ และไบอัส b มีค่าเริ่มต้นเท่ากับ 1.0 ค่าเอาต์พุตของเครือข่ายคือ

$$ y = \mathsf{hardlim}(w_{11}p_1 + w_{12}p2 + b) \tag{9.21}$$

โดยที่ $\mathbf{p} = [p_1, p_2]^T$ ค่าของ ${}_1\mathbf{w}$ และ b จะถูกปรับด้วยกฎการเรียนรู้ของเพอร์เซ็พตรอน - II (ตามความสัมพันธ์ที่ (9.13) ข้างต้น) รายละเอียดเชิงตัวเลขมีแสดงในตารางที่ 9.1 ค่าเอาต์พุตของเครือข่ายจะถูกคำนวณจากเวกเตอร์อินพุตแต่ละตัว และจะนำเอาไปเปรียบเทียบกับเวกเตอร์เป้าหมายของอินพุตนั้นๆ จากตารางดังกล่าว จะเห็นได้ว่าการเรียนรู้ของเครือข่ายหยุดลง เมื่อทำการคำนวณหาเอาต์พุตจากเวกเตอร์อินพุตทุกตัว แล้วค่าเอาต์พุตที่ได้เป็นไปตามเวกเตอร์เป้าหมายทุกๆ เวกเตอร์อินพุต (นั่นคือค่าความผิดพลาด $e = 0$) รูปที่ 9.7 แสดงการปรับตัวของค่าน้ำหนักประสาท ${}_1\mathbf{w}$ และไบอัส b ในระหว่างการเรียนรู้ของเครือข่าย



รูปที่ 9.6: เวกเตอร์อินพุต และ เวกเตอร์เป้าหมาย พร้อม เส้นแบ่ง พื้นที่ เริ่มต้น โดยที่ P(1) และ P(2) แทนองค์ประกอบของอินพุตเวกเตอร์ $\mathbf{p} = [p_1, p_2]^T$

จากรูปที่ 9.7 จะเห็นได้ว่าการฝึกสอนเครือข่ายเพอร์เซ็พตรอนจะขึ้นอยู่กับค่าเริ่มต้นของน้ำหนักประสาทและไบอัสด้วย ถ้าเส้นแบ่งพื้นที่ (หรือค่าน้ำหนักประสาท) อยู่ใกล้กับคำตอบ การเรียนรู้ของเครือข่ายก็จะใช้เวลาเร็วขึ้นในการค้นหาเส้นแบ่งพื้นที่ที่ต้องการ ■

ตัวอย่างข้างต้นแสดงให้เห็นถึงการทำงานและกระบวนการเรียนรู้ของเครือข่ายเพอร์เซ็พตรอน เราสามารถพิจารณาการฝึกสอนเครือข่ายประสาทเทียม ว่าเป็นการค้นหาค่าน้ำหนักประสาทและไบอัสที่เหมาะที่สุดของเครือข่ายได้

9.3 ข้อจำกัดของเครือข่ายเพอร์เซ็พตรอน

การเรียนรู้ของเครือข่ายเพอร์เซ็พตรอนได้รับการพิสูจน์ว่าลู่เข้าสู่คำตอบได้ในจำนวนครั้งที่จำกัด (ไม่เป็นอนันต์) ตราบเท่าที่คำตอบนั้นมีอยู่จริง อย่างไรก็ตามเครือข่ายเพอร์เซ็พตรอนไม่สามารถใช้ในการแก้ปัญหาบางอย่างได้ ในกรณีที่ผ่านมาเพอร์เซ็พตรอนแบบนิวรอนเดียวสามารถคัดแยกพื้นที่อินพุตออกเป็น 2 พื้นที่ ซึ่งขอบเขตการแบ่งนิยามด้วยเส้นแบ่งพื้นที่

$$ {}_1\mathbf{w}\mathbf{p} + b = 0 \tag{9.22} $$

จะเห็นได้ชัดว่าเส้นแบ่งพื้นที่ดังกล่าวเป็นเส้นตรง (ระนาบหลายมิติ) ดังนั้นเพอร์เซ็พตรอนสามารถใช้กับเวกเตอร์อินพุตที่แบ่งได้ด้วยเส้นตรงเท่านั้น เราเรียกเส้นแบ่งพื้นที่ดังกล่าวว่าเป็นเส้น**แบ่งพื้นที่ได้แบบเชิงเส้น** (linearly separable) ตัวอย่างเช่นในปัญหาการกระทำ AND อย่างไรก็ตามปัญหาหลายๆ อย่างเป็นแบบ**แบ่งพื้นที่ได้แบบ**

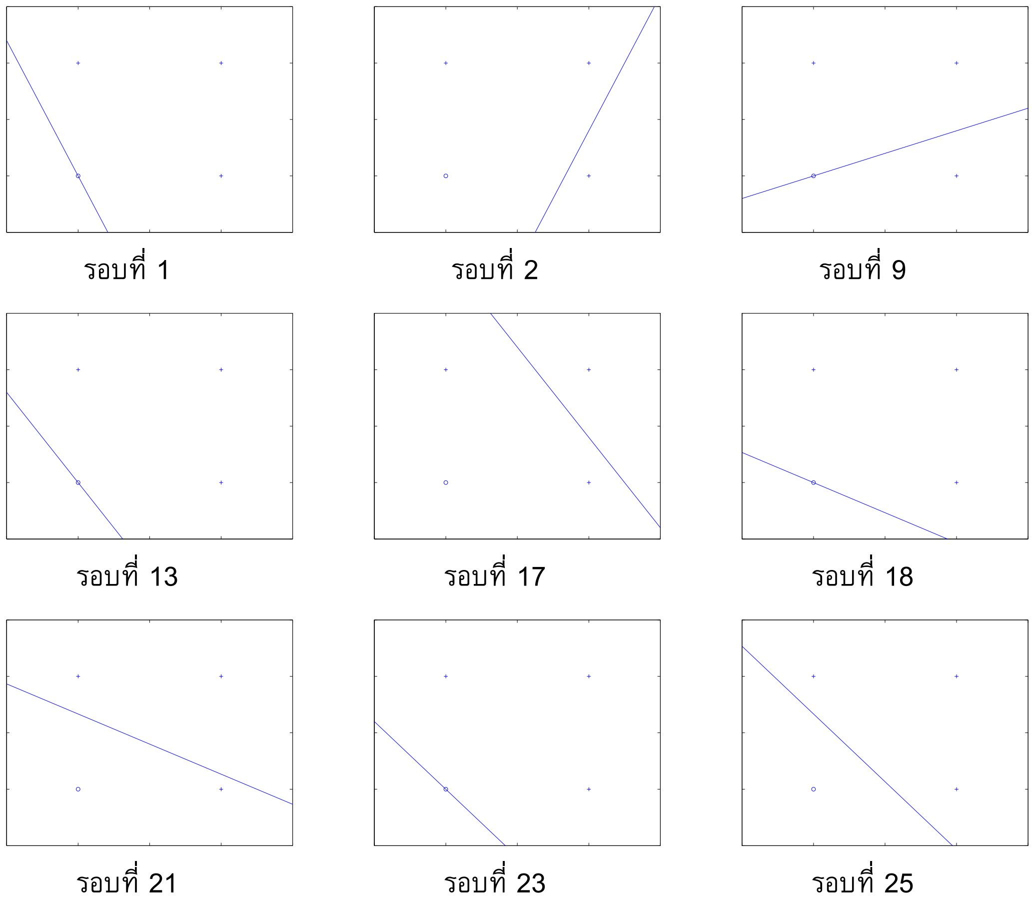
รอบที่	$\mathbf{p}$	$\mathbf{y}$	$\mathbf{t}$	e	w_{11}	w_{12}	b
-	-	-	-	-	-1.2	-0.5	+1.0
1	$\mathbf{p_1}$	+1	0	-1	-1.2	-0.5	0.0
2	$\mathbf{p_2}$	0	+1	+1	-1.2	+0.5	+1.0
3	$\mathbf{p_3}$	0	+1	+1	-0.2	+0.5	+2.0
4	$\mathbf{p_4}$	+1	+1	0	-0.2	+0.5	+2.0
5	$\mathbf{p_1}$	+1	0	-1	-0.2	+0.5	+1.0
6	$\mathbf{p_2}$	+1	+1	0	-0.2	+0.5	+1.0
7	$\mathbf{p_3}$	+1	+1	0	-0.2	+0.5	+1.0
8	$\mathbf{p_4}$	+1	+1	0	-0.2	+0.5	+1.0
9	$\mathbf{p_1}$	+1	0	-1	-0.2	+0.5	0.0
10	$\mathbf{p_2}$	+1	+1	0	-0.2	+0.5	0.0
11	$\mathbf{p_3}$	0	+1	+1	+0.8	+0.5	+1.0
12	$\mathbf{p_4}$	+1	+1	0	+0.8	+0.5	+1.0
13	$\mathbf{p_1}$	+1	0	-1	+0.8	+0.5	0.0
14	$\mathbf{p_2}$	+1	+1	0	+0.8	+0.5	0.0
15	$\mathbf{p_3}$	+1	1	0	+0.8	+0.5	0.0
16	$\mathbf{p_4}$	+1	+1	0	+0.8	+0.5	0.0
17	$\mathbf{p_1}$	+1	0	-1	+0.8	+0.5	-1.0
18	$\mathbf{p_2}$	0	+1	+1	+0.8	+1.5	0.0
19	$\mathbf{p_3}$	+1	+1	0	+0.8	+1.5	0.0
20	$\mathbf{p_4}$	+1	+1	0	+0.8	+1.5	0.0
21	$\mathbf{p_1}$	+1	0	-1	+0.8	+1.5	-1.0
22	$\mathbf{p_2}$	+1	+1	0	+0.8	+1.5	-1.0
23	$\mathbf{p_3}$	0	+1	+1	+1.8	+1.5	0.0
24	$\mathbf{p_4}$	+1	+1	0	+1.8	+1.5	0.0
25	$\mathbf{p_1}$	+1	0	-1	+1.8	+1.5	-1.0
26	$\mathbf{p_2}$	+1	+1	0	+1.8	+1.5	-1.0
27	$\mathbf{p_3}$	+1	+1	0	+1.8	+1.5	-1.0
28	$\mathbf{p_4}$	+1	+1	0	+1.8	+1.5	-1.0
29	$\mathbf{p_1}$	0	0	0	+1.8	+1.5	-1.0

ตารางที่ 9.1: การปรับค่าน้ำหนักประสาทและไบอัสขณะเรียนรู้ของเครือข่ายเพอร์เซ็พตรอน

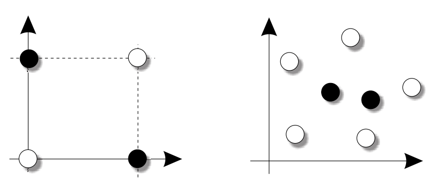
ไม่เป็นเชิงเส้น (linearly inseparable) ยกตัวอย่างเช่นปัญหาอมตะ XOR หรือตัวอย่างในรูปที่ 9.8 อย่างไรก็ตาม ได้มีการนำเสนอเครือข่ายเพอร์เซ็พตรอนแบบหลายชั้นสำหรับปัญหาแบบแบ่งพื้นที่ได้แบบไม่เป็นเชิงเส้น ซึ่งจะได้กล่าวถึงในรายละเอียดภายหลัง

9.4 การปรับแต่งกฎการเรียนรู้แบบเพอร์เซ็พตรอน

กฎการเรียนรู้แบบเพอร์เซ็พตรอนได้รับการปรับปรุงหลากหลายวิธี Mays [Widrow and Lehr, 1990] ได้เสนอวิธีการปรับปรุงกฎการเรียนรู้ดั้งเดิมของเพอร์เซ็พตรอน 2 วิธี ทั้งสองวิธีเพิ่มการพิจารณาใช้*พื้นที่เฉพาะ* (dead zone) ในการเลือกการปรับค่าน้ำหนักประสาท พื้นที่เฉพาะดังกล่าวมีรัศมี γ ถ้าเน็ตเอาต์พุต n มีขนาดน้อยกว่า γ



รูปที่ 9.7: การปรับตัวของเวกเตอร์น้ำหนักประสาทและไบอัสระหว่างการเรียนรู้ เส้นแบ่งพื้นที่มีการเปลี่ยนแปลงตามค่าของ w_{11} w_{12} และ b ที่เปลี่ยนแปลงตามกฎการเรียนรู้ของเพอร์เซ็พตรอนในแต่ละรอบการคำนวณ



รูปที่ 9.8: ตัวอย่างปัญหาแบบแบ่งพื้นที่ได้แบบไม่เป็นเชิงเส้น

แสดงว่าเน็ตเอาต์พุตอยู่ในพื้นที่เฉพาะนี้ (นั่นคือถ้า $|n| < \gamma$) อัลกอริทึมในการปรับค่าน้ำหนักประสาทสามารถสรุปได้ดังนี้

▷ **อัลกอริทึมการปรับส่วนเพิ่ม (increment adaptation algorithm)**

$$\mathbf{w}^{new} = \begin{cases} \mathbf{w}^{old} + \alpha e \frac{\mathbf{p}}{\|\mathbf{p}\|^2} & \text{ถ้า } |n| \geq \gamma \\ \mathbf{w}^{old} + \alpha t \frac{\mathbf{p}}{\|\mathbf{p}\|^2} & \text{ถ้า } |n| < \gamma \end{cases} \tag{9.23}$$

โดยที่ α คือค่าคงที่การเรียนรู้ ถ้าพื้นที่เฉพาะนี้เป็นศูนย์ กฎการเรียนรู้ในสมการ (9.23) จะเป็นกฎการเรียนรู้แบบเพอร์เซ็พตรอนทั่วๆ ไปเหมือนในสมการ (9.19) Mays ได้พิสูจน์กฎการเรียนรู้ข้างต้นแล้วว่าในกรณีแบ่งแยกได้เชิงเส้น (linearly separable) อัลกอริทึมดังกล่าวจะลู่เข้าสู่คำตอบเสมอ และสามารถที่จะแยกแยะรูปแบบได้ในจำนวนการเรียนรู้ที่จำกัด ในกรณีแบ่งแยกได้แบบไม่เป็นเชิงเส้น (nonlinearly separable) อัลกอริทึมนี้ยังมีประสิทธิภาพดีกว่ากฎการเรียนรู้ดั้งเดิมแบบเพอร์เซ็พตรอนอันเนื่องมาจากพื้นที่เฉพาะ ถ้าพื้นที่เฉพาะมีขนาดใหญ่เพียงพอ การปรับเวกเตอร์น้ำหนักประสาทจะมีทิศทางที่ไม่เป็นเวกเตอร์ศูนย์ถ้ามีคำตอบที่ดีพออยู่จริง พื้นที่เฉพาะที่ใหญ่เพียงพอนี้ยังมีผลให้ค่าความผิดพลาดเฉลี่ยที่ได้มีค่าต่ำ

สำหรับกฎการเรียนรู้แบบเพอร์เซ็พตรอนดั้งเดิม อินพุตที่มีรูปแบบที่ไม่สามารถแยกได้แบบเชิงเส้นจะนำไปสู่จำนวนขั้นตอนการเรียนรู้ที่ไม่สิ้นสุด และโดยปกติแล้วจะให้ค่าความผิดพลาดเฉลี่ยที่มีค่าสูง Mays ยังได้พิสูจน์อีกว่าการใช้พื้นที่เฉพาะทำให้ลดความไวของค่าความผิดพลาดในน้ำหนักประสาทได้อีกด้วย

▷ **อัลกอริทึมปรับปรุงการผ่อนคลาย (modified relaxation)**

$$\mathbf{w}^{new} = \begin{cases} \mathbf{w}^{old} & \text{ถ้า } |n| \geq \gamma \text{ และ } e = 0 \\ \mathbf{w}^{old} + \alpha \hat{e} \frac{\mathbf{p}}{\|\mathbf{p}\|^2} & \text{อื่นๆ} \end{cases} \tag{9.24}$$

โดยที่ $\hat{e} = y - n$ ในกรณีที่พื้นที่เฉพาะมีขนาดเป็นอนันต์ ($\gamma \to \infty$) อัลกอริทึมข้างต้นจะกลายเป็นกฎการเรียนรู้แบบเพอร์เซ็พตรอนดั้งเดิม ถ้าพื้นที่เฉพาะเป็น $0 < \gamma < 1$ และค่าคงที่การเรียนรู้ $0 < \alpha \leq 2$ อัลกอริทึมปรับปรุงการผ่อนคลายนี้จะลู่เข้าเสมอ และสามารถแยกแยะรูปแบบของอินพุต ที่เป็นแบบแยกแยะได้เชิงเส้นในจำนวนขั้นตอนการเรียนรู้ที่จำกัด แต่ถ้าอินพุตเป็นแบบแยกแยะได้แบบไม่เป็นเชิงเส้น อัลกอริทึมนี้จะทำงานเหมือนกับอัลกอริทึมการปรับส่วนเพิ่ม

9.5 สรุป

เนื้อหาในบทนี้ได้นำเสนอกฎการเรียนรู้แรก ที่เราได้ศึกษาสำหรับเครือข่ายประสาทเทียม อันได้แก่กฎการเรียนรู้แบบเพอร์เซ็พตรอน ซึ่งเป็นการเรียนรู้แบบมีผู้ฝึกสอน กระบวนการฝึกสอนเครือข่ายเพอร์เซ็พตรอนเริ่มจากการป้อนคู่ตัวอย่างอินพุต/เอาต์พุต (หรือเป้าหมาย) ให้กับเครือข่าย เครือข่ายจะทำการปรับค่าพารามิเตอร์ตามกฎการเรียนรู้ เพื่อให้เอาต์พุตของเครือข่ายเคลื่อนที่เข้าสู่เป้าหมายที่กำหนดไว้ในคู่ตัวอย่างอินพุต/เอาต์พุต

กฎการเรียนรู้แบบเพอร์เซ็พตรอนเป็นกฎที่ไม่ซับซ้อน แต่ทรงประสิทธิภาพ กฎดังกล่าวจะลู่เข้าสู่คำตอบเสมอ (ถ้าคำตอบมีจริง) อย่างไรก็ดี เครือข่ายเพอร์เซ็พตรอนก็มีข้อจำกัด ที่ซึ่งไม่สามารถแก้ปัญหาที่แบ่งแยกได้แบบไม่เป็นเชิงเส้นได้ ข้อจำกัดดังกล่าวถูกแก้ไขด้วยเครือข่ายที่มีประสิทธิภาพมากกว่าเพอร์เซ็พตรอน อันได้แก่เครือข่ายไปข้างหน้าแบบหลายชั้น รวมไปถึงกฎการเรียนรู้ที่ทรงประสิทธิภาพกว่ากฎการเรียนรู้แบบเพอร์เซ็พตรอน ดังจะได้กล่าวถึงในหัวข้อต่อไป

โจทย์คำถาม

9.1. ออกแบบเครือข่ายเพอร์เซ็พตรอนในการคัดแยกอินพุตเวกเตอร์ที่มี 4 กลุ่มดังนี้

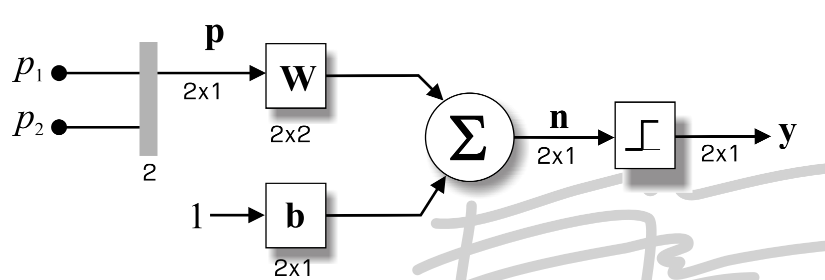
- กลุ่มที่ 1 $\left\{\mathbf{p}_1 = \begin{bmatrix} -0.5 \\ 2 \end{bmatrix}, \mathbf{p}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right\}$
- กลุ่มที่ 2 $\left\{\mathbf{p}_3 = \begin{bmatrix} 1.5 \\ -0.5 \end{bmatrix}, \mathbf{p}_4 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}\right\}$
- กลุ่มที่ 3 $\left\{\mathbf{p}_5 = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \mathbf{p}_6 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}\right\}$
- กลุ่มที่ 4 $\left\{\mathbf{p}_7 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \mathbf{p}_8 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$

เครือข่ายเพอร์เซ็พตรอนต้องออกแบบให้มีอย่างน้อย 2 นิวรอน (สำหรับ 4 กลุ่ม) โดยใช้โครงสร้างของเครือข่ายในรูปที่ 9.9 อินพุตเวกเตอร์ซึ่งแบ่งเป็น 4 กลุ่มพร้อมตัวอย่างเส้นแบ่งพื้นที่ 2 เส้นแสดงในรูปที่ 9.10 เวกเตอร์เป้าหมายสามารถกำหนดได้ดังนี้

- กลุ่มที่ 1 $\left\{\mathbf{t}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \mathbf{t}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}\right\}$
- กลุ่มที่ 2 $\left\{\mathbf{t}_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \mathbf{t}_4 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$
- กลุ่มที่ 3 $\left\{\mathbf{t}_5 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{t}_6 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$
- กลุ่มที่ 4 $\left\{\mathbf{t}_7 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{t}_8 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$

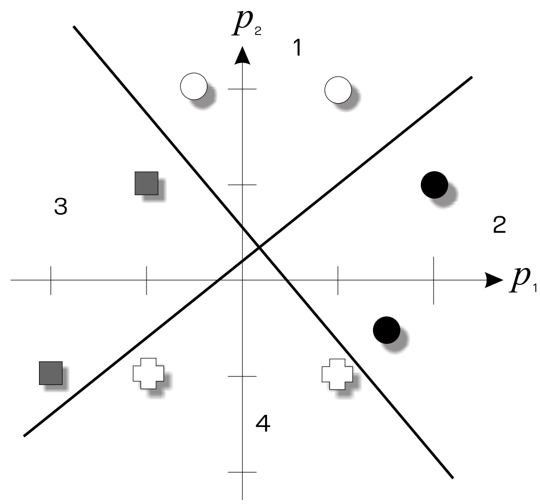
จากรายละเอียดข้างต้น ให้ตอบคำถามต่อไปนี้

- ให้ออกแบบคำนวณหาค่าเมตริกซ์น้ำหนักประสาทและไบอัส (มีได้หลายคำตอบ) แสดงรายละเอียดการออกแบบและคำนวณ วิเคราะห์และสรุปผลที่ได้
- ปรับปรุงการเรียนรู้ของเพอร์เซ็พตรอนโดยใช้วิธีของ Mays ทดลองเลือกค่าพารามิเตอร์ต่างๆ และวิเคราะห์ผลลัพธ์ที่ได้ โดยเปรียบเทียบกับการเรียนรู้เพอร์เซ็พตรอนแบบดั้งเดิม

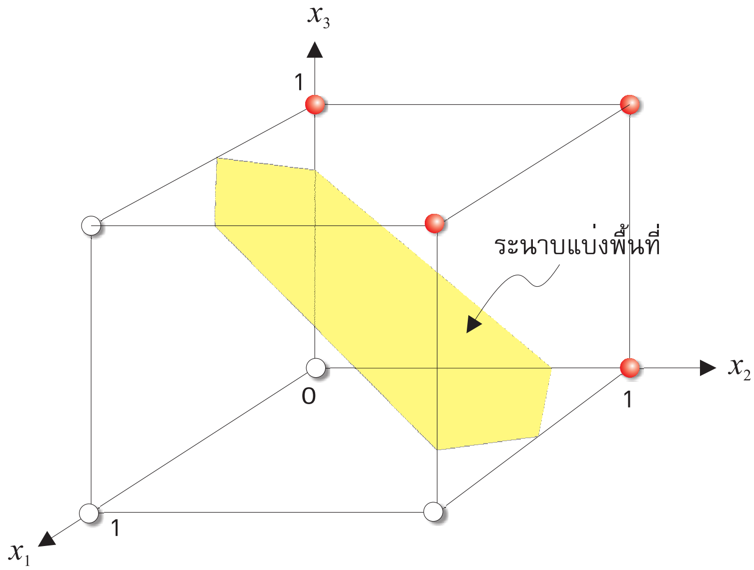


รูปที่ 9.9: เพอร์เซ็พตรอนแบบ 2 นิวรอน

9.2. เพอร์เซ็พตรอนสามารถใช้เป็นตัวกระทำการทางลอจิกได้ จงออกแบบเครือข่ายเพอร์เซ็พตรอนที่ทำหน้าที่ฟังก์ชันลอจิกแบบไบนารี AND OR NOT และ COMPLEMENT อภิปรายพารามิเตอร์ในการออกแบบและผลลัพธ์ที่ได้



รูปที่ 9.10: อินพุตเวกเตอร์และเส้นแบ่งพื้นที่

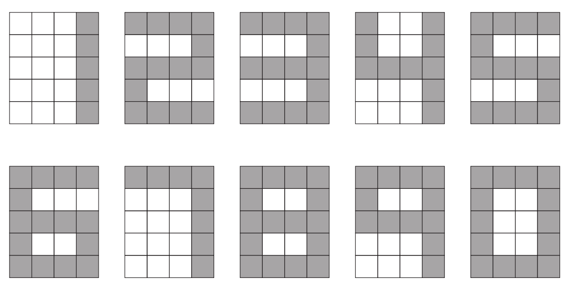


รูปที่ 9.11: ข้อมูล 2 กลุ่มในระนาบ 3 มิติ

9.3. รูปที่ 9.11 แสดงข้อมูล 2 กลุ่มในระนาบ 3 มิติ จงออกแบบเครือข่ายเพอร์เซ็พตรอนในการแบ่งกลุ่มข้อมูลในรูป พร้อมแสดงรายละเอียดการฝึกสอนและอภิปรายผลลัพธ์ที่ได้

9.4. จงทดลองออกแบบเครือข่ายเพอร์เซ็พตรอนสำหรับฟังก์ชันลอจิก XOR แบบ 2 มิติ อภิปรายผลลัพธ์ที่ได้ (*หมายเหตุ* เครือข่ายเพอร์เซ็พตรอนไม่สามารถใช้เป็นฟังก์ชันลอจิก XOR ได้)

9.5. เพอร์เซ็พตรอนสามารถใช้เป็นตัวคัดแยกเชิงเส้น (linear classifier) ได้ จงพิจารณาออกแบบเครือข่ายเพอร์เซ็พตรอนสำหรับคัดแยกตัวเลขในรูปที่ 9.12 โดยจำนวนของนิวรอนในเครือข่ายควรจะต้องมีขนาดเท่ากับจำนวนหลักของตัวเลข แต่ละหลักแทนด้วยเมตริกซ์ขนาด 5×4 (ทำการแปลงให้อยู่ในรูปเวกเตอร์ก่อนที่จะป้อนให้กับเครือข่าย) เอาต์พุตของแต่ละนิวรอนจะสัมพันธ์กับหลักของตัวเลข เช่นเอาต์พุตของนิวรอนตัวที่ 1 ควรจะตอบสนองกับอินพุตที่เป็นเลข 1 เมื่อทำการฝึกสอนและทดสอบเครือข่ายแล้วให้ทำการใส่สัญญาณรบกวนให้กับอินพุต แล้วทำการทดสอบกับเครือข่ายอีกครั้ง สังเกตประสิทธิภาพของเครือข่ายพร้อมทั้งอภิปรายผลลัพธ์ที่ได้



รูปที่ 9.12: เวกเตอร์ตัวเลข 10 หลัก

บรรณานุกรม

W.S. McCulloch and W. Pitts. A logical calculus of ideas immanent in nervous activity. In *Bulletin of Mathematical Biophysics*, volume 5, pages 115--133, 1943.

F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. In *Psychological Review*, volume 65, pages 386--408, 1958.

F. Rosenblatt. On the convergence of reinforcement procedures in simple perceptrons. Report, Cornell Aeronautical Lab, February 1960.

F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Washington: Spartan Books, 1962.

B. Widrow and M.A. Lehr. 30 years of neural networks: Perceptron, madaline and backpropagation. In *Proceedings of the IEEE*, volume 78, pages 1415--1442, 1990.