บทที่ 3

วิธีดำเนินการวิจัย

ในการดำเนินการวิจัยสำหรับการสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์เพื่อทำนายค่าออกซิเจนในน้ำผิวดิน กรณีศึกษาคุณภาพน้ำคลองในเขตดุสิต ประกอบด้วยขั้นตอนดังต่อไปนี้

3.1 การรวบรวมข้อมูลปริมาณค่าพารามิเตอร์ที่มีผลต่อคุณภาพน้ำในเขตดุสิต

ทำการเก็บรวบรวมข้อมูลปริมาณคุณภาพน้ำจากการบันทึกข้อมูลของเจ้าหน้าที่สำนักการระบายน้ำกรุงเทพตั้งแต่ปี 2549 – 2551 รวม 4 ปี จำนวน 11 จุดเก็บตัวอย่างซึ่งเป็นจุดเก็บตัวอย่างที่อยู่ในเขตดุสิต คือ จุดเก็บที่ 74 (หน้ากรมวิเทศสหการ) จุดเก็บที่ 75 (ตลาดเทวราช) จุดเก็บที่ 76 (ประตูรับน้ำเทเวศร์) จุดเก็บที่ 81 (ประตูรับน้ำสามเสน) จุดเก็บที่ 82 (วัดโบสถ์) จุดเก็บที่ 112 (สี่แยกสะพานแดง) จุดเก็บที่ 116 (ถนนเศรษฐศิริ) จุดเก็บที่ 117 (ถนนศรีอยุธยา) จุดเก็บที่ 121 (สะพานพิบูลสงคราม) จุดเก็บที่ 741 (ถนนสามเสน บริเวณบริษัท บุญรอด บริวเวอรี่) และจุดเก็บที่ 801 (วัดราชาธิวาช) โดยแต่ละจุดเก็บจะประกอบด้วยข้อมูลพารามิเตอร์ที่มีผลต่อคุณภาพน้ำที่จัดเก็บเป็นรายเดือน ดังนี้ ค่าความเป็นกรด-ด่าง (pH) ปริมาณจุลินทรีย์ที่ใช้ในการสลายสารอินทรีย์ในน้ำ (BOD) ปริมาณออกซิเจนที่สารเคมีใช้ในการย่อยสลายอินทรีย์ในน้ำ (COD) ปริมาณของแข็งในน้ำ (SS) ปริมาณไนโตรเจนที่มีอยู่ในน้ำเสีย (TKN) ปริมาณแอมโมเนีย-ไนโตรเจน (NH_3N), ปริมาณไนไตร-ไนโตรเจน (NO_2N), ปริมาณไนเตรท-ไนโตรเจน (NO_3N), ปริมาณฟอสฟอรัสทั้งหมด (T-P) and ปริมาณแบคทีเรียกลุ่มโคลิฟอร์มทั้งหมด (total coliform) จะกำหนดให้เป็นตัวแปรอิสระและข้อมูลปริมาณค่าออกซิเจนในน้ำ (DO) ให้เป็นตัวแปรตาม หลังจากนั้นนำค่าตัวแปรต่าง ๆ มาสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์

3.2 การศึกษาสัมประสิทธิ์สหสัมพันธ์

ในการศึกษาสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่างตัวแปรต่าง ๆ กับปริมาณค่าออกซิเจนในน้ำ จะใช้เทคนิควิธีทางสถิติในการวิเคราะห์เพื่อหาความสัมพันธ์ของตัวแปรว่ามีความสัมพันธ์กันมากหรือน้อย และเป็นการทดสอบว่าตัวแปรมีความสัมพันธ์กันในรูปเชิงเส้นหรือไม่ รวมถึงการทดสอบว่าตัวแปรมีความสัมพันธ์กันในทิศทางเดียวกันหรือไม่ โดยจะใช้สัญลักษณ์ R แทนสัมประสิทธิ์สหสัมพันธ์ ซึ่งถ้าผลการศึกษาที่ได้พบว่า R เป็นลบ แสดงว่าตัวแปรที่ศึกษามีความสัมพันธ์ในทิศทางตรงข้ามกัน แต่ถ้า R เป็นบวก แสดงว่าตัวแปรมีความสัมพันธ์ในทิศทางเดียวกัน และถ้า R มีค่าเข้าใกล้ 1 (ไม่พิจารณาที่เครื่องหมาย) แสดงว่าตัวแปรที่ศึกษามีความสัมพันธ์กันมาก และในทางกลับกันถ้า R มีค่าเข้าใกล้ 0 แสดงว่าตัวแปรมีความสัมพันธ์กันน้อยแต่ถ้าหาก R มีค่า เท่ากับ 0 แสดงว่า ตัวแปรไม่มีความสัมพันธ์กันเลย โดยค่า R สามารถหาได้จากสมการที่ (3.1)

$$R = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\left(\sum(X-\bar{X})^2\sum(Y-\bar{Y})^2\right)}} \qquad (3.1)$$

ในการศึกษาครั้งนี้ผู้วิจัยได้แยกกรณีศึกษาสัมประสิทธิ์สหสัมพันธ์ของตัวแปรตาม และตัวแปรอิสระ โดยกำหนดให้มีจำนวนตัวแปรอิสระ 1 ช่วงเวลาจัดเก็บ (โดยมีการจัดเก็บข้อมูลเดือนละหนึ่งครั้ง) ดังตารางที่ 3.1

ตารางที่ 3.1 รูปแบบข้อมูลป้อนเข้า

	ตัวแปรอิสระ	ตัวแปรตาม
เวลาปัจจุบัน	$(pH)_t$ $(BOD)_t$ $(COD)_t$ $(SS)_t$ $(TKN)_t$ $(NH_3N)_t$ $(NO_2N)_t$ $(NO_3N)_t$ $(T\text{-}P)_t$ และ $(\text{total coliform})_t$	DO_{t+1}
เวลาย้อนหลัง 1 ช่วงเวลา	$(pH)_{t-1}$ $(BOD)_{t-1}$ $(COD)_{t-1}$ $(SS)_{t-1}$ $(TKN)_{t-1}$ $(NH_3N)_{t-1}$ $(NO_2N)_{t-1}$ $(NO_3N)_{t-1}$ $(T\text{-}P)_{t-1}$ และ $(\text{total coliform})_{t-1}$	

โดยจะทำการศึกษาสัมประสิทธิ์สหสัมพันธ์ต่อโดยแยกออกเป็น 4 กรณีย่อย ดังนี้

1) ตัวแปรอิสระ กับ $(DO)_{t+1}$
2) Log ของตัวแปรอิสระ กับ $(DO)_{t+1}$
3) ตัวแปรอิสระ กับ Log ของ $(DO)_{t+1}$
4) Log ของตัวแปรอิสระ กับ Log ของ $(DO)_{t+1}$

เมื่อทำการศึกษาสัมประสิทธิ์สหสัมพันธ์ของกรณีศึกษาแบบต่าง ๆ แล้ว จะทำการคัดเลือกตัวแปรที่เหมาะสมเพื่อนำไปใช้ในการสร้างแบบจำลองต่อไป

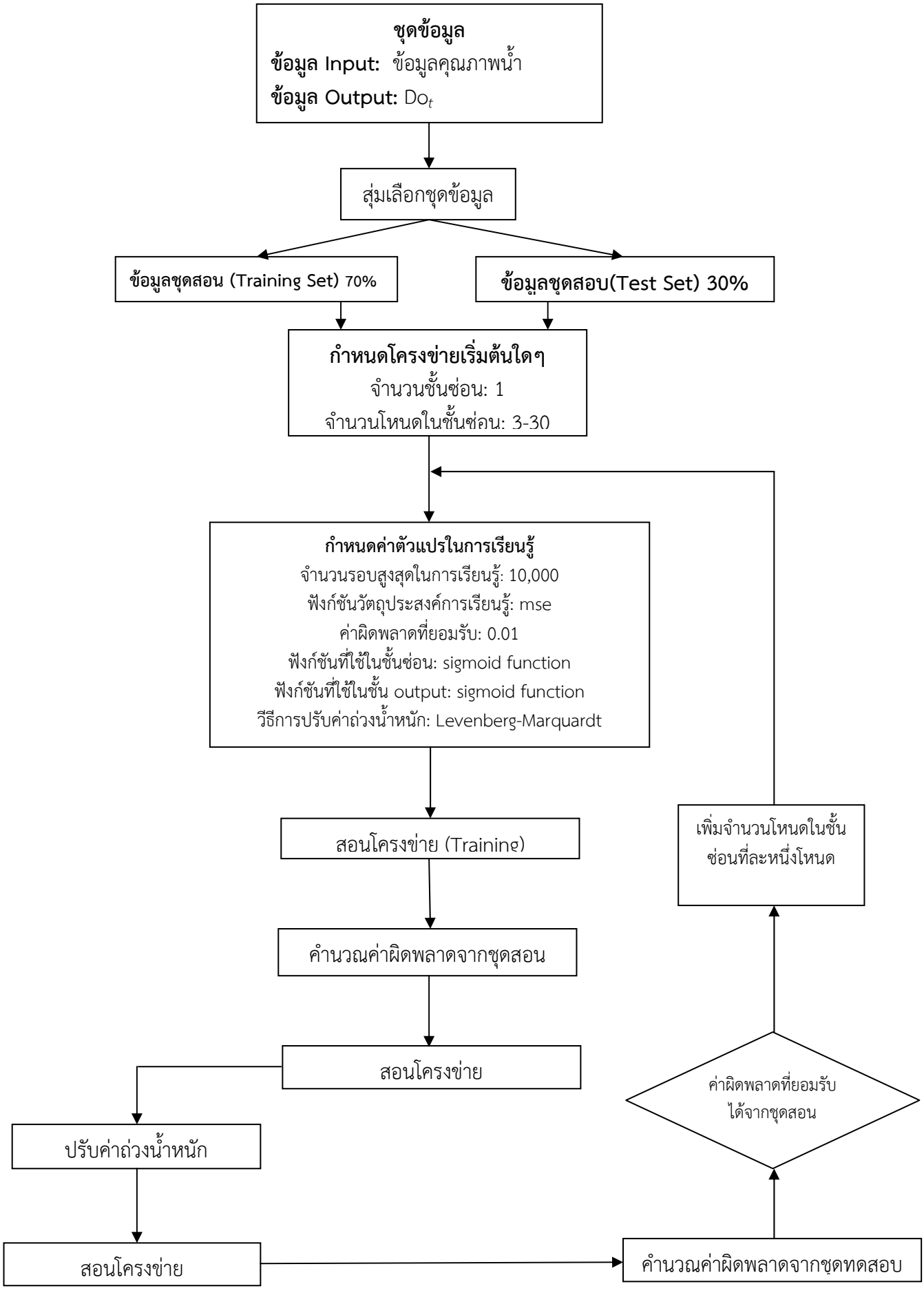
3.3 การสร้างแบบจำลองโครงข่ายใยประสาทประดิษฐ์

การประมาณค่าออกซิเจนในน้ำโดยใช้โครงข่ายใยประสาทประดิษฐ์ประกอบด้วยขั้นตอนใหญ่ ๆ 4 ขั้นตอน คือ การเตรียมข้อมูล การออกแบบโครงข่าย การสอนโครงข่าย และการตรวจสอบความถูกต้องของโครงข่าย ดังแสดงในหน้าถัดไป (Jiang et al., 2004) โดยเตรียมข้อมูลจากจุดเก็บตัวอย่าง ซึ่งจะกำหนดให้ค่าออกซิเจนในน้ำ (DO) เป็นตัวแปรตาม และค่าพารามิเตอร์ตัวอื่นๆอีก 10

ตัวที่บ่งบอกคุณภาพน้ำเป็นตัวแปรอิสระ โดยเรียงข้อมูลตามเดือน เป็นจำนวน 4 ปี หรือ 384 ชุดข้อมูล (Patterns) หลังจากนั้นจะแบ่งชุดข้อมูลออกเป็น 2 ชุด คือข้อมูลชุดสอน (training set) และข้อมูลชุดทดสอบ (test set) โดยแบ่งออกเป็นร้อยละ 70 และ 30 ตามลำดับ หลังจากนั้นทำการออกแบบโครงข่ายโดยกำหนดให้เป็นแบบไปข้างหน้าหลายชั้น (MLFF) ซึ่งเป็นระบบแบบมีครูสอน โดยจะใช้วิธีการสอนแบบแพร่กระจายความผิดพลาดกลับ (Error back – propagation algorithm) จากนั้นนำข้อมูลชุดสอน (training set) ให้โครงข่ายได้เรียนรู้ โครงข่ายจะประมวลผลจนได้คำตอบชุดหนึ่ง สำหรับคำตอบที่โครงข่ายสามารถคำนวณออกมาได้นั้น จะถูกนำมาหาค่าความผิดพลาดโดยเปรียบเทียบกับค่าออกซิเจนในน้ำ (DO) จริงจากจุดเก็บตัวอย่างว่ามีค่ามากน้อยเพียงใด ถ้ายังมีความผิดพลาดสูงอยู่ ระบบจะย้อนกลับไปปรับเปลี่ยนค่าถ่วงน้ำหนัก และทำการสอนต่อไปจนกว่าค่าความผิดพลาดระหว่างคำตอบที่ได้จากโครงข่าย และคำตอบจริง จะมีค่าน้อยในระดับที่ยอมรับได้ซึ่งในขณะเดียวกันนั้น ชุดทดสอบ (test set) จะทำการทดสอบค่าความผิดพลาดของโครงข่ายไปพร้อมๆกัน เมื่อค่าความผิดพลาดจากชุดทดสอบมีค่าน้อยในระดับที่ยอมรับได้ จึงจะหยุดทำการปรับสอนและได้โครงข่ายที่เหมาะสมสำหรับใช้งาน ซึ่งรายละเอียดของขั้นตอนการสร้างโครงข่ายใยประสาทประดิษฐ์ในงานวิจัยครั้งนี้ มีดังต่อไปนี้

3.3.1 การเตรียมข้อมูล

ในการศึกษาครั้งนี้มีพื้นที่กรณีศึกษา คือ ในเขตดุสิตโดยแต่ละพื้นที่จะนำข้อมูลคุณภาพน้ำจากจุดเก็บตัวอย่างของแต่ละแห่งมาหาความสัมพันธ์กับค่าออกซิเจน (DO) ในเดือนถัดไป โดยกำหนดให้ปริมาณค่าออกซิเจนในน้ำ (DO) เป็นตัวแปรตาม และและค่าพารามิเตอร์ตัวอื่นๆอีก 10 ตัวที่บ่งบอกคุณภาพน้ำเป็นตัวแปรอิสระ โดยจะทำการศึกษาหาความสัมพันธ์ภายในจุดเก็บเดียวกัน ซึ่งจะทำการศึกษาสัมประสิทธิ์สหสัมพันธ์ และคัดเลือกตัวแปรที่เข้าสู่โครงข่ายโดยพิจารณาจากที่กล่าวไว้ในหัวข้อที่ 3.2



รูปที่ 3.1 ขั้นตอนการสร้างโครงข่ายใยประสาทประดิษฐ์

3.3.2 การออกแบบโครงข่าย

การออกแบบโครงข่ายจะพิจารณาจากจำนวนข้อมูลป้อนเข้าโครงข่าย จำนวนชั้นซ่อน จำนวนโหนดในชั้นซ่อน และจำนวนผลลัพธ์เป็นหลัก ซึ่งจะต้องทำการหาค่าให้เหมาะสมกับโครงข่ายมากที่สุด โดยมีรายละเอียด ดังนี้

1) ข้อมูลป้อนเข้าที่เหมาะสม

ข้อมูลป้อนเข้าต้องพิจารณาหลาย ๆ แบบด้วยกัน เนื่องจากเป็นข้อมูลทางสิ่งแวดล้อมที่ส่วนใหญ่มีการกระจายตัวแบบ log - normal โดยแบ่งกรณีศึกษาดังที่กล่าวไว้แล้วในหัวข้อที่ 3.2

2) การหาจำนวนชั้นซ่อนที่เหมาะสม

โครงข่ายใยประสาทประดิษฐ์ที่ใช้ในการศึกษาครั้งนี้ ซึ่งข้อมูลของพารามิเตอร์ที่ใช้เป็นข้อมูลคุณภาพน้ำที่มีความซับซ้อน จะใช้แบบ ระบบโครงข่ายใยประสาทประดิษฐ์ที่มี 3 ชั้น

- ระบบโครงข่ายใยประสาทประดิษฐ์ที่มี 3 ชั้น

ระบบโครงข่ายใยประสาทประดิษฐ์ที่มี 3 ชั้น ประกอบไปด้วยชั้นข้อมูลป้อนเข้า 1 ชั้น (1 input layer) ชั้นซ่อน 1 ชั้น (1 hidden layer) และชั้นผลลัพธ์ 1 ชั้น (1 output layer) ดังรูปที่ 3.2

โดยที่

$$\Sigma = \sum_{i=1}^{n} w_{ij} x_i = \left(net_{ij}\right)_h \qquad (3.2)$$

$$f = \frac{1}{1+\exp\left(\left(-net_{ij}\right)_h\right)} = f\left(\left(net_{ij}\right)_h\right) \qquad (3.3)$$

w_{ij} = ค่าถ่วงน้ำหนักชั้นซ่อน (hidden layer)

w_{jk} = ค่าถ่วงน้ำหนักชั้นผลลัพธ์ (output layer)

h = ชั้นซ่อน (hidden layer)

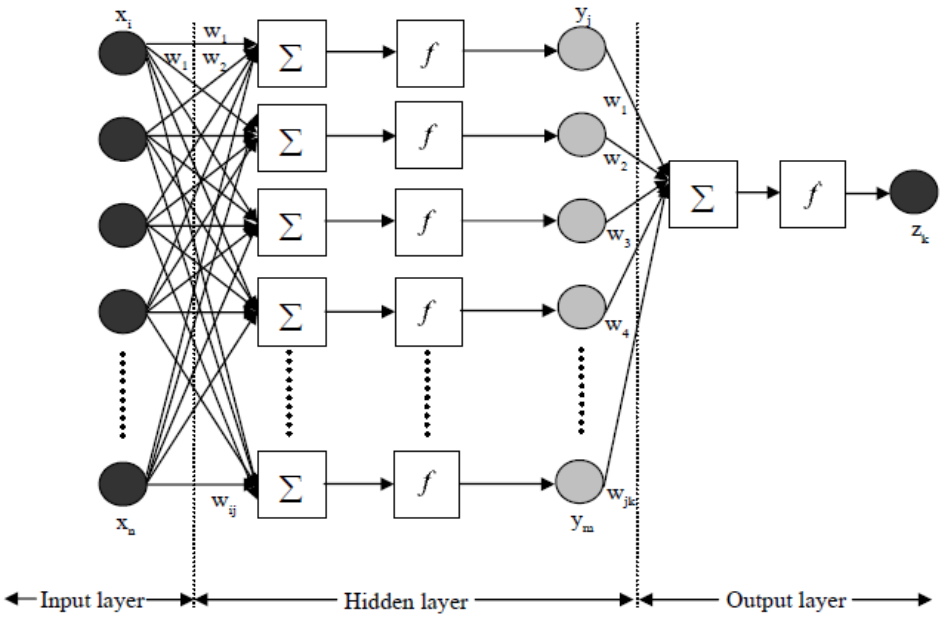
o = ชั้นผลลัพธ์ (output layer)

x_i = ข้อมูลป้อนเข้า (input data) คือ ค่าพารามิเตอร์ที่บ่งบอกคุณภาพน้ำ ได้แก่ (pH) (BOD) (COD) (SS) (TKN) (NH_3N) (NO_2N) (NO_3N) (T-P) และ (total coliform)ณ เดือนใด ๆ

Y_j = ผลลัพธ์ของชั้นซ่อน คือ ผลลัพธ์ที่ได้จากชั้นซ่อนชั้นสุดท้าย

Z_k = ผลลัพธ์ (output data) คือ ค่าออกซิเจนในน้ำ (DO) ที่โครงข่ายคำนวณได้

T_k = ค่าจริง (actual data) คือ ค่าออกซิเจนในน้ำ (DO) ที่ได้จากจุดตรวจวัดจริง



รูปที่ 3.2 โครงข่ายใยประสาทประดิษฐ์ที่มี 3 ชั้น

3) การหาจำนวนโหนดในชั้นซ่อนที่เหมาะสม

ในการกำหนดจำนวนโหนดในชั้นซ่อน ไม่มีกฎเกณฑ์ หรือทฤษฎีที่แน่นอนเพราะเมื่อกำหนดจำนวนโหนดในชั้นซ่อนมาก จะทำให้เสียเวลาในการสอนมาก เนื่องจากจำนวนการเชื่อมต่อของแต่ละโหนดมีจำนวนมาก แต่ในทางกลับกัน ถ้ากำหนดให้มีจำนวนโหนดในชั้นซ่อนที่น้อยจนเกินไป โครงข่ายอาจจะไม่สามารถเรียนรู้จนพบคำตอบที่แท้จริงได้ ซึ่งในการศึกษาครั้งนี้จะกำหนดจำนวนโหนดในชั้นซ่อนเริ่มต้นตามค่าแนะนำของโปรแกรม (MATLAB R2009a) คือ

จำนวนโหนดในชั้นซ่อน = 1/2 (Inputs + Outputs) + รากที่สองของจำนวนของ patterns ที่ใช้ในการสอน และใช้ค่าจำนวนโหนดที่ปรับจากค่าแนะนำ อีก 4 ค่า คือ Default+10 Default+20 Default-10 Default-20 ยกตัวอย่างการกำหนดจำนวนโหนดในชั้นซ่อน อาทิเช่น ชุดข้อมูลป้อนเข้า (Inputs) รวมตัวแปรทั้งหมด 1 สถานี มีจำนวน 10 ตัวแปร ส่วนผลลัพธ์ (Output) ที่ต้องการ คือ ค่าค่าออกซิเจนในน้ำ (DO) ล่วงหน้า มีจำนวน 1 ตัวแปร และจำนวน patterns ที่ใช้ในการสอนจากร้อยละ 70 ของจำนวน patterns ทั้งหมด คือ 384 x 70% = 269 patterns ดังนั้น จำนวนโหนดในชั้นซ่อน คือ 1/2 (10+1) +√384 = 25 โหนด และใช้ค่าจำนวนโหนดตามค่าดีฟอลท์ที่ปรับอีก 4 ค่า คือ 5, 15, 35 และ 45 โหนด โดยประมาณ

4) การหาค่าพารามิเตอร์ที่เหมาะสม (อัตราการเรียนรู้ โมเมนตัม ค่าถ่วงน้ำหนัก)

การเลือกค่าถ่วงน้ำหนัก และค่าโมเมนตัมจะมีอิทธิพลต่อการเกิดค่า error ของ

โครงข่ายได้ โดยค่าถ่วงน้ำหนักไม่ควรเป็นค่าที่ใหญ่มากนัก จะเป็นผลทำให้ค่าอนุพันธ์ของฟังก์ชั่น การแปลงค่าแบบฟังก์ชันซิกมอยด์มีค่าเล็กมาก เรียกว่า อยู่ในย่านของการอิ่มตัว (saturation Region) แต่ถ้าค่าน้ำหนักมีค่าเล็กเกินไป จะทำให้ค่าที่จะส่งไปยังโหนดในชั้นซ่อน หรือโหนดในชั้นผลลัพธ์จะมีค่าเข้าใกล้ศูนย์ ซึ่งเป็นสาเหตุทำให้การเรียนรู้ทำได้ช้า โดยทั่วไปค่าถ่วงน้ำหนักเริ่มต้นจะสุ่มค่าระหว่าง -0.5 ถึง 0.5 (หรือ ระหว่าง -1 ถึง 1 ตามความเหมาะสม) (สุรยุทธ ปรัชญา, 2541) ซึ่งในการศึกษาครั้งนี้ จะกำหนดค่าพารามิเตอร์เริ่มต้น และช่วงการปรับเปลี่ยนค่าโดยปรับครั้งละ 0.1 ดังตารางที่ 3.2

ตารางที่ 3.2 ค่าพารามิเตอร์ในโครงข่าย

พารามิเตอร์	ค่าเริ่มต้น	ช่วงการปรับเปลี่ยน
อัตราการเรียนรู้	0.05	0.05, 0.1, 0.2
ค่าถ่วงน้ำหนัก	0.2	0.2, 0.4, 0.6, 0.8
โมเมนตัม	0.1	0.1, 0.3, 0.5, 0.7

3.3.3 การสอนโครงข่าย

- วิธีการสอน

ในขั้นตอนการสอนโครงข่ายนี้จะใช้วิธีการแพร่กระจายความผิดพลาดกลับ (Error back –propagation algorithm) โดยใช้ฟังก์ชันซิกมอยด์เป็นฟังก์ชั่นการแปลงค่า (Transfer Function) เนื่องจากฟังก์ชั่นการแปลงค่ามีความสำคัญมากในโครงข่ายแบบแพร่กระจายความผิดพลาดกลับ ซึ่งฟังก์ชั่นที่ใช้ควรมีความต่อเนื่อง ไม่เป็นเชิงเส้น สามารถหาค่าอนุพันธ์ได้ และง่ายต่อการคำนวณ ซึ่งค่าอนุพันธ์สามารถเขียนในรูปเทอมของฟังก์ชันนั้น โดยฟังก์ชันซิกมอยด์มีรูปแบบสมการดังสมการที่ (3.4)

$$f(x) = \frac{1}{1+\exp(-x)} \qquad (3.4)$$

- การหยุดการสอน

ในแต่ละรอบของการสอน โครงข่ายจะทำการตรวจสอบค่าผิดพลาดด้วยชุดทดสอบไปพร้อมกัน เพื่อแก้ปัญหาการเกิด “over fitting” ซึ่งหมายถึง โครงข่ายที่มีรอบการเรียนรู้ที่สูงเกินไปทำให้เกิดการจดจำค่าของชุดข้อมูล แต่จะไม่มีการเรียนรู้ โดยจะให้ผลทดสอบที่สูงในข้อมูลชุดสอน แต่เมื่อนำชุดทดสอบมาตรวจสอบจะให้ผลทดสอบที่ต่ำ โดยในการศึกษาครั้งนี้จะตั้งข้อกำหนดในการ

หยุดการสอน เมื่อโครงข่ายเรียนรู้ได้ 10,000 รอบ หรือค่าความผิดพลาดที่ชุดทดสอบมีค่าเท่ากับ 0.01

3.3.4 การตรวจสอบความถูกต้องของโครงข่าย

การประเมินผลจากการทำนายโดยใช้โครงข่ายใยประสาทประดิษฐ์ สามารถใช้สถิติในการทดสอบความถูกต้อง ซึ่งสถิติที่ใช้ในการวัดความผิดพลาดระหว่างผลการทำนายและข้อมูลจริงจากชุดตรวจสอบความถูกต้อง (Validate set) คือ ค่าความผิดพลาดกำลังสองเฉลี่ย (Mean square error, MSE) และค่าร้อยละความผิดพลาดเฉลี่ย (Mean absolute error, MAE)

1) ค่าความผิดพลาดกำลังสองเฉลี่ย (Mean square error, MSE) สามารถคำนวณได้ดังสมการที่ (3.5)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Predict_i - Actual_i)^2 \qquad (3.5)$$

2) ค่าร้อยละความผิดพลาดเฉลี่ย (Mean absolute error, MAE) สามารถคำนวณได้ดังสมการที่ (3.6)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\frac{Predict_i - Actual_i}{Actual_i} \times 100 \qquad (3.6)$$

โดยที่ Actual = ข้อมูลจริงที่ได้จากจุดตรวจวัด

Predict = ข้อมูลที่ได้จากการทำนายด้วยโครงข่ายใยประสาทประดิษฐ์

N = จำนวนชุดข้อมูลที่ใช้ในการคำนวณ

โดยใช้เกณฑ์การพิจารณายอมรับโครงข่าย คือ ต้องเป็นโครงข่ายที่มีลำดับที่ (rank) ที่น้อยที่สุด โดยนำค่า MAE ของกรณีที่พิจารณาหนึ่ง ๆมาทดสอบความแตกต่างของ ผลของชั้นซ่อนและจำนวนโหนดในชั้นซ่อนที่เหมาะสม