บทที่ 4
การปรับอัตราการเรียนรู้บนข่ายงานระบบประสาทแบบชั้นเดียว

ในบทนี้ แสดงการทดสอบประสิทธิภาพของวิธีปรับอัตราการเรียนรู้แบบต่างๆ เปรียบเทียบกับการใช้อัตราการเรียนรู้เป็นค่าคงตัว ตัวอย่างที่เลือกใช้ในบทนี้เป็นฟังก์ชันเชิงเส้น เพราะว่าเป็นปัญหาที่ง่ายและชัดเจน ทำให้สามารถวิเคราะห์ผลของวิธีต่างๆ ได้ในเบื้องต้น ข้อมูลตัวอย่างสำหรับการทดสอบ สร้างจากฟังก์ชันที่กำหนดไว้ ซึ่งจะมีทั้งกรณีที่มีการเปลี่ยนแปลงตามเวลา (stationary) และไม่มีการเปลี่ยนแปลงตามเวลา (nonstationary)

การทดสอบได้เลือกใช้ข่ายงานระบบประสาทแบบชั้นเดียวและเลือกฟังก์ชัน activation เป็นแบบเชิงเส้น การทดสอบทำโดยให้ระบบทำการประมาณฟังก์ชัน (function approximation) ฟังก์ชันที่ใช้อ้างอิงเพื่อทดสอบ จะกำหนดให้มีโดเมน (domain) ในช่วงเดียวกับสัญญาณขาเข้า $\underline{x}(t)$ และสัญญาณอ้างอิง $d(t)$ ก็สามารถกำหนดได้จากฟังก์ชันอ้างอิงนี้ ฟังก์ชันอ้างอิงมีลักษณะเป็นเชิงเส้น ทำให้ระบบสามารถประมาณได้อย่างแม่นยำ ดังนั้น ความผิดพลาดที่วัดได้จะไม่มีผลจากข้อจำกัดของระบบรวมอยู่ด้วย

4.1 การวัดประสิทธิภาพการเรียนรู้ของข่ายงานระบบประสาทแบบชั้นเดียว

การวัดประสิทธิภาพใช้เป็นผลรวมความผิดพลาดกำลังสอง ความผิดพลาดสามารถวัดได้จากผลต่างระหว่างสัญญาณอ้างอิง $d(t)$ และสัญญาณขาออก $o(t)$ ดังนั้น ฟังก์ชันสำหรับการวัดประสิทธิภาพ $P(\underline{x}(t))$ จะเป็นไปตามสมการ

$$P(\underline{x}(t)) = \int_{a_n}^{b_n} \cdots \int_{a_2}^{b_2} \int_{a_1}^{b_1} \left[\left(d(t) - o(t) \right)^2 \right] dx_1 dx_2 \cdots dx_n \qquad (4.1.1)$$

โดยที่ x_i เป็นส่วนประกอบที่ i ของเวกเตอร์สัญญาณขาเข้า $\underline{x}$ ซึ่งมีจำนวน N สัญญาณ ส่วน a_i และ b_i เป็นขอบล่าง (lower bound) และขอบบน (upper bound) ของ x_i ตามลำดับ การทดสอบในบทนี้ ได้ใช้ขอบล่างและขอบบนของ x_i เป็น -0.5 และ 0.5 ตามลำดับ

4.2 ฟังก์ชันอ้างอิงของการทดสอบ

ฟังก์ชันอ้างอิงที่ใช้เป็นแบบเชิงเส้น ซึ่งมีรูปทั่วไป ดังนี้

$$f(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{N} a_i x_i + K \qquad (4.2.1)$$

โดยที่ $f(x_1, x_2, \ldots, x_N)$ เป็นผลรวมถ่วงน้ำหนักของตัวแปรอิสระ x_i, $1 \le i \le N$ และค่าคงตัวค่า a_i, $1 \le i \le N$ และ K เป็นค่าคงตัว ซึ่งเป็นสัมประสิทธิ์และเทอมค่าคงตัวของฟังก์ชัน ตามลำดับ สมการ (4.2.1) นี้เป็นรูปแบบฟังก์ชันที่ใช้สำหรับหาสัญญาณอ้างอิง $d(t)$ จากค่าสัญญาณขาเข้า $\underline{x}(t)$ ซึ่งจำนวนเทอมจะขึ้นกับการกำหนดจำนวนของตัวแปรอิสระ

4.3 การทดสอบกับฟังก์ชันอ้างอิงที่ไม่มีการเปลี่ยนแปลงตามเวลา (stationary)

ในหัวข้อนี้ จะแสดงผลและการเปรียบเทียบเบื้องต้นของการทดสอบในกรณีที่ฟังก์ชันอ้างอิงไม่เปลี่ยนแปลงตามเวลา โดยใช้วิธีการปรับอัตราการเรียนรู้ ดังนี้

- ใช้อัตราการเรียนรู้เป็นค่าคงตัว ที่ 0.01
- ปรับอัตราการเรียนรู้ตามวิธี Dbd โดยกำหนดค่าคงตัวสำหรับการเพิ่ม κ เป็น 0.00035, ค่าคงตัวสำหรับการลด φ เป็น 0.003 และค่าถ่วงน้ำหนักสำหรับอดีต θ เป็น 0.7
- ปรับอัตราการเรียนรู้ตามวิธี Idbd โดยกำหนดค่า meta step size (η) เป็น 0.0001
- ปรับอัตราการเรียนรู้ตามวิธี Entropy โดยกำหนดค่าถ่วงน้ำหนักสำหรับสมการหาค่าเฉลี่ยตามเวลา θ เป็น 0.6
- ปรับอัตราการเรียนรู้ตามวิธี direction cosine ของ Franzini [14]
- ปรับอัตราการเรียนรู้ตามวิธี modified direction cosine

วิธีการปรับอัตราการเรียนรู้ทั้ง 5 แบบ จะถูกทดสอบประสิทธิภาพกับฟังก์ชันอ้างอิงที่มีตัวแปรอิสระแตกต่างกัน 3 ค่า คือ 3 ตัวแปร, 5 ตัวแปร และ 10 ตัวแปร ซึ่งมีรูปสมการ ดังนี้

- ฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 3 ตัว จะใช้สมการในรูป

$$y = x_1 + x_2 + x_3 \qquad (4.3.1)$$

- ฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 5 ตัว จะใช้สมการในรูป

$$y = x_1 + x_2 + x_3 + x_4 + x_5 \qquad (4.3.2)$$

- ฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 10 ตัว จะใช้สมการในรูป

$$y = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 + x_{10} \qquad (4.3.3)$$

โดยค่าตัวแปรอิสระทุกตัวในฟังก์ชันอ้างอิงที่นำมาใช้ทดสอบ จะกระจายอย่างสม่ำเสมอ ในช่วง [-0.5, 0.5]

รูปที่ 4.2 ถึง 4.4 แสดงผลการทดสอบการใช้อัตราการเรียนรู้ทั้ง 4 แบบ กับฟังก์ชันอ้างอิงมีตัวแปรอิสระ 3, 5 และ 10 ตัวแปร ตามลำดับ ฟังก์ชันไม่มีการเปลี่ยนแปลงตามเวลา การทดสอบได้กำหนดค่าเริ่มต้นสำหรับแต่ละวิธี (ใช้ N แทนจำนวนสัญญาณขาเข้า) ดังนี้

- วิธี Dbd กำหนดค่าเริ่มต้นของอัตราการเรียนรู้ $\beta_i(0),\ 1 \le i \le N$ เป็น 0.01 และการเปลี่ยนแปลงพารามิเตอร์ $\delta_i(0),\ 1 \le i \le N$ เป็น 0

- วิธี Idbd กำหนดค่าเริ่มต้นของพารามิเตอร์เสริม $\alpha_i(0),\ 1 \le i \le N$ เป็น $\log\left(\frac{1}{N}\right)$ และ $h_i(0),\ 1 \le i \le N$ เป็น 0 และค่าเริ่มต้นของอัตราการเรียนรู้ ซึ่งเป็นไปตามขั้นตอนวิธี $\beta_i(0) = \exp\left(\alpha_i\right),\ 1 \le i \le N$

- วิธี Entropy กำหนดค่าเริ่มต้นของอัตราการเรียนรู้ $\beta_i(0),\ 1 \le i \le N$ เป็น 0.01 , ค่าเฉลี่ยพารามิเตอร์ $\overline{w}_i(0),\ 1 \le i \le N$ เท่ากับ 0 และค่าเริ่มต้นของความแปรปรวน $v_i(0),\ 1 \le i \le N$ เป็น 0.1

- วิธี direction cosine กำหนดค่าเริ่มต้นของอัตราการเรียนรู้เป็น 0.01

- วิธี modified direction cosine กำหนดค่าเริ่มต้นของอัตราการเรียนรู้เป็น 0.01

และค่าเริ่มต้นของทุกพารามิเตอร์ได้จากการสุ่มค่าในช่วง [-0.5, 0.5] การทดสอบทำโดยสุ่มเลือกค่าตัวแปรอิสระจากทั้งหมด 10,000 ค่า เพื่อเป็นสัญญาณขาเข้า ส่วนค่าสัญญาณอ้างอิง $d(t)$ หาได้จากสมการ (4.4.1) การทดสอบแต่ละกรณีจะให้ระบบเรียนรู้ทั้งหมด 20 ครั้ง แต่ละครั้งมีลำดับสัญญาณขาเข้าต่างกัน ผลจากการทดสอบที่ได้นำมาเฉลี่ยกันเพื่อให้มีความถูกต้องมากขึ้น และสรุปผลหลังจากเรียนรู้ 1, 5, 10 และ 20 ครั้ง ทุกวิธีใช้ค่าเริ่มต้นเดียวกันเพื่อให้สามารถนำผลการทดสอบมาเปรียบเทียบกันได้ การวัดประสิทธิภาพที่แสดงในรูป จะเป็นผลรวมความผิดพลาดกำลังสองในช่วงโดเมนของสัญญาณขาเข้าคือ [-0.5, 0.5] ดังแสดงในสมการ (4.2.1)

ผลการทดสอบแสดงเป็นเส้นกราฟ โดยแต่ละเส้นจะใช้แทนผลจากวิธีการปรับอัตราการเรียนรู้ ดังนี้

- อัตราการเรียนรู้คงตัวที่ 0.01 เป็นเส้นทึบ ดังนี้

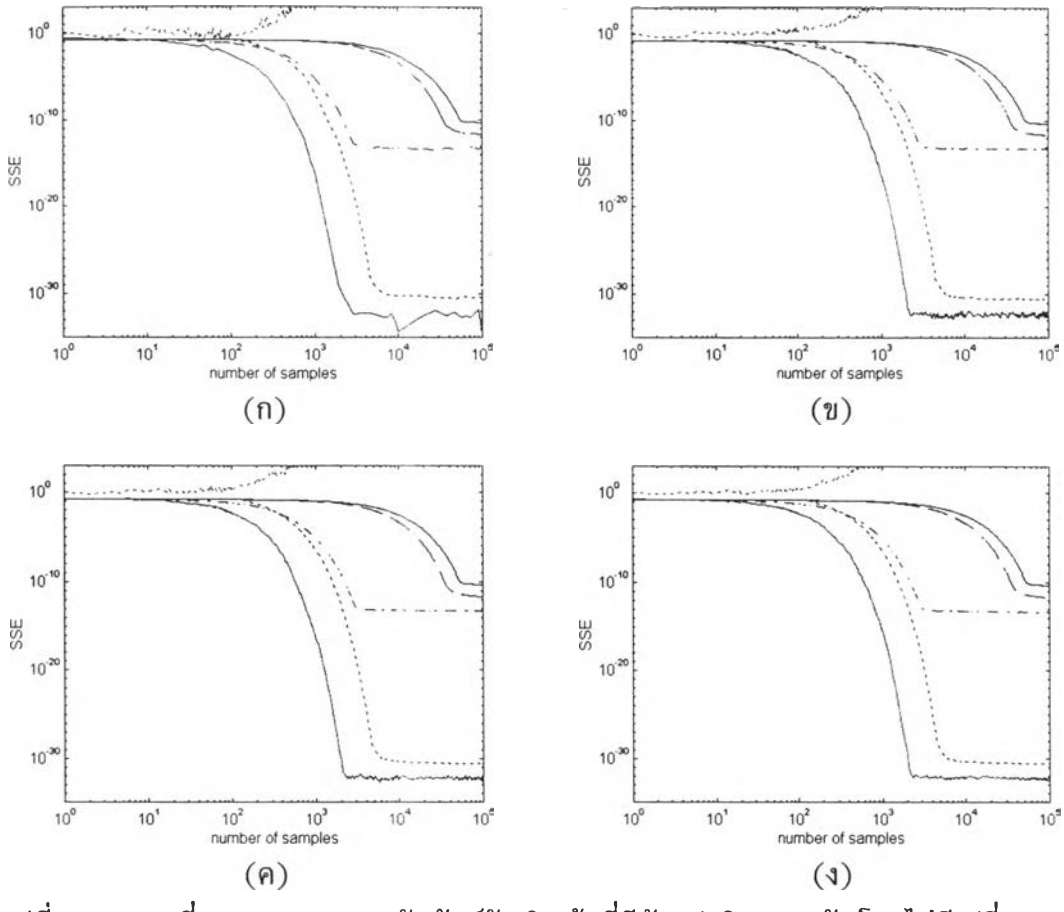
- วิธี Dbd ใช้เส้นประที่มีขนาดยาว ดังนี้

- วิธี Idbd ใช้เส้นประสลับจุด ดังนี้

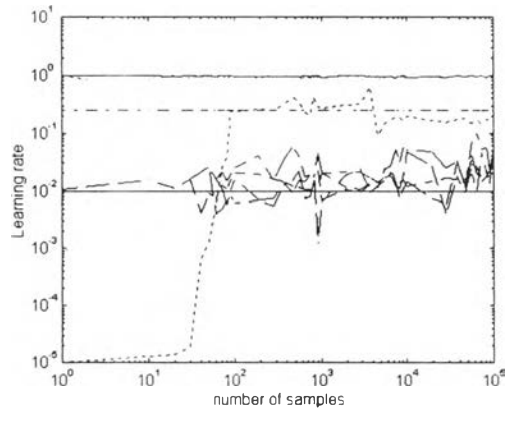
- วิธี Entropy ใช้เส้นที่เป็นจุด ดังนี้

- วิธี direction cosine ที่มีการปรับปรุงให้อัตราการเรียนรู้เป็นบวกเสมอเป็นเส้นทึบหนา ดังนี้ (กราฟ SSE เป็นเส้นบนที่เพิ่มขึ้นเสมอ)

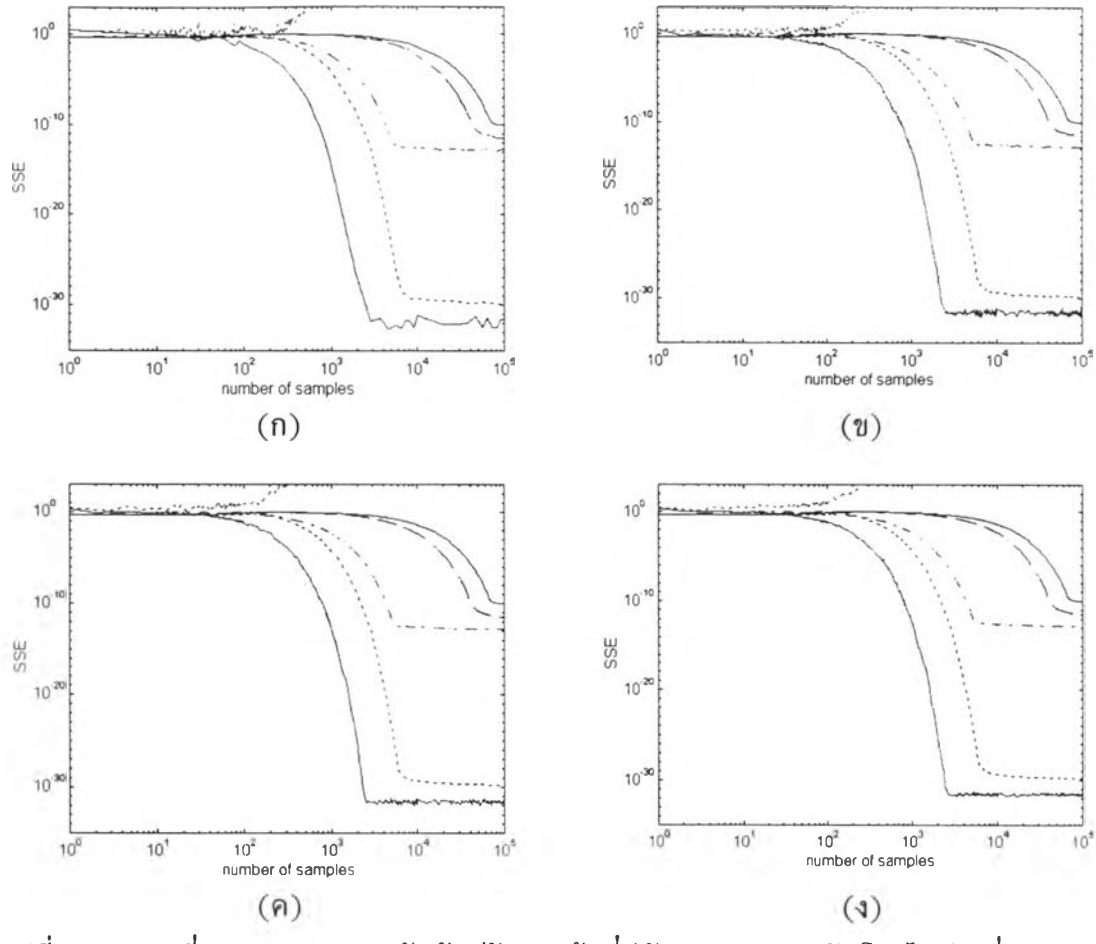
- วิธี modified direction cosine ที่มีการปรับปรุงให้อัตราการเรียนรู้เป็นบวกเสมอเป็นเส้นประหนา ดังนี้ (กราฟ SSE เป็นเส้นต่ำสุด)



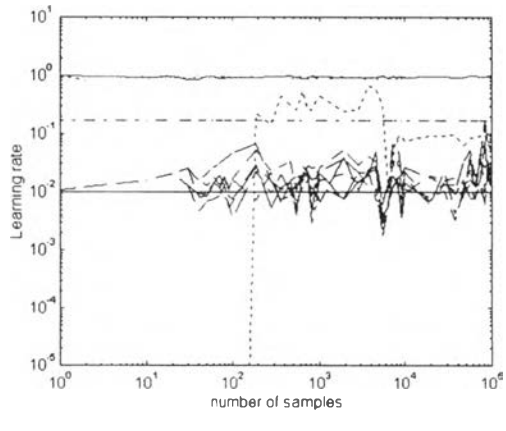
รูปที่ 4.1 ผลเฉลี่ยจากการทดสอบกับฟังก์ชันเชิงเส้นที่มีตัวแปรอิสระ 3 ตัว โดยไม่มีเปลี่ยนแปลงตามเวลา (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



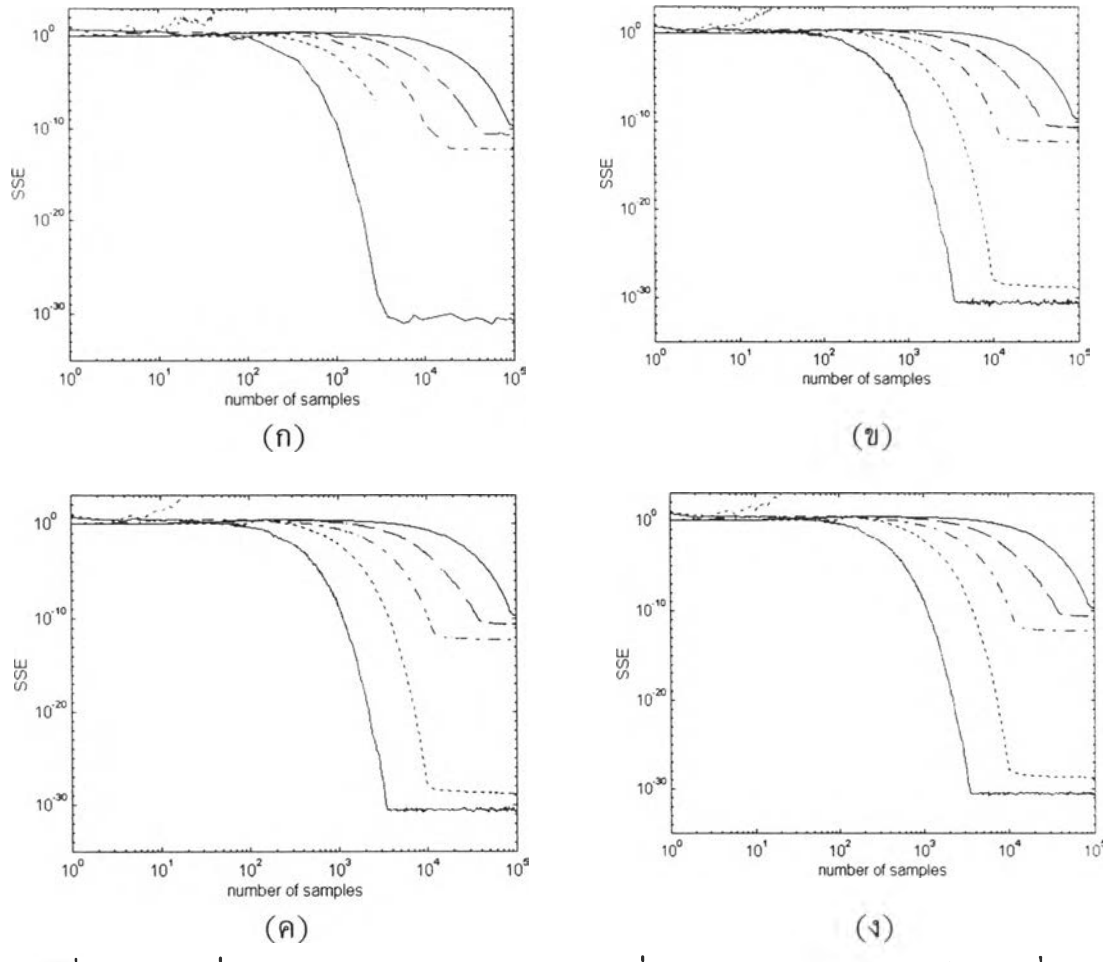
รูปที่ 4.2 อัตราการเรียนรู้ในกรณีที่ใช้ฟังก์ชันอ้างอิงมีตัวแปรอิสระ 3 ตัว



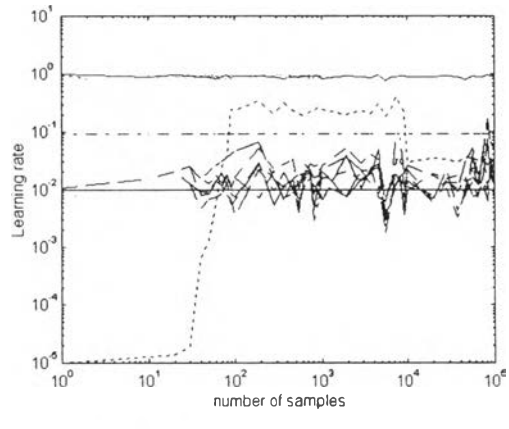
รูปที่ 4.3 ผลเฉลี่ยจากการทดสอบกับฟังก์ชันเชิงเส้นที่มีตัวแปรอิสระ 5 ตัว โดยไม่มีเปลี่ยนแปลงตามเวลา (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.4 อัตราการเรียนรู้ในกรณีที่ใช้ฟังก์ชันอ้างอิงมีตัวแปรอิสระ 5 ตัว



รูปที่ 4.5 ผลเฉลี่ยจากการทดสอบกับฟังก์ชันเชิงเส้นที่มีตัวแปรอิสระ 10 ตัว โดยไม่มีเปลี่ยนแปลงตามเวลา (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.6 อัตราการเรียนรู้ในกรณีที่ใช้ฟังก์ชันอ้างอิงมีตัวแปรอิสระ 10 ตัว

รูปที่ 4.1 แสดงผลการเรียนรู้ในกรณีที่ใช้ฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 3 ตัว เพื่อทดสอบการใช้อัตราการเรียนรู้ทั้ง 5 แบบ โดยแสดงกราฟของผลรวมความผิดพลาดกำลังสอง (sum-square error) หรือ SSE เทียบกับจำนวนของข้อมูลที่ป้อนให้ระบบ (number of samples) ผลการทดสอบแสดงให้เห็นว่า วิธี modified direction cosine สามารถทำให้ระบบปรับตัวได้ดีที่สุด โดยสามารถเข้าสู่สถานะอยู่ตัวได้เร็วและมีผลรวมความผิดพลาดกำลังสองต่ำสุด ถัดมาจึงเป็นวิธี Entropy ซึ่งสามารถให้ค่าที่สถานะอยู่ได้ใกล้เคียงกัน แต่เข้าสู่สถานะอยู่ช้ากว่า วิธี direction cosine ทำให้ระบบไม่ลู่เข้าจากอัตราการเรียนรู้ที่เป็นลบ ส่วนวิธีอื่น ๆ ให้ค่าในสถานะอยู่ตัวที่สูงกว่ามาก และเข้าได้ช้ากว่า โดยการปรับอัตราการเรียนรู้ทุกวิธีสามารถให้ค่าผลรวมกำลังสองได้ต่ำกว่าและเข้าสู่สถานะอยู่ตัวได้เร็วกว่าการใช้อัตราการเรียนรู้เป็นค่าคงตัวที่ 0.01

เมื่อพิจารณาประกอบกับค่าอัตราการเรียนรู้ในรูปที่ 4.2 เห็นได้ว่า อัตราการเรียนรู้ของวิธี modified direction cosine มีค่าสูงสุด คือ อยู่ใกล้ 1 ตลอดการปรับตัวของระบบ ส่วนอัตราการเรียนรู้ของวิธี Entropy มีค่ามากขึ้นหลังจากได้รับข้อมูล 100 ชุด จนมีค่ามากที่สุดในช่วง 200 ถึง 2,000 ชุด แล้วจึงมีค่าลดลงเล็กน้อย ในขณะที่วิธี Idbd ให้ค่าใกล้เคียงกันตลอดการเรียนรู้ และวิธี Dbd ให้ค่าอัตราการเรียนรู้ที่แกว่งไปมา และเพิ่มขึ้นอย่างช้า ๆ กราฟ SSE ของวิธี Entropy ในรูปที่ 4.1 สอดคล้องกับอัตราการเรียนรู้ที่เปลี่ยนแปลงนี้ โดยค่า SSE เริ่มมีค่าลดลงอย่างรวดเร็วหลังจากได้รับข้อมูล 100 ชุด และเพิ่มขึ้นเรื่อย ๆ อัตราการเรียนรู้ในช่วงนี้จะมีค่าสูงสุด จนกระทั่ง เมื่อระบบเข้าสู่สถานะอยู่ตัว อัตราการเรียนรู้ก็ลดลง ทำให้ระบบปรับตัวเข้าใกล้ค่าเหมาะสมที่สุดได้มากขึ้น

วิธี Dbd ให้อัตราการเรียนรู้ที่มีค่าสูงกว่า 0.01 เล็กน้อย ส่วนวิธี Idbd ให้อัตราการเรียนรู้ที่ค่อนข้างคงตัวที่มีค่าสูงกว่า กราฟ SSE ของทั้งสองวิธี จึงลดลงเร็วกว่ากราฟของอัตราการเรียนรู้คงตัว วิธี Idbd จะลดลงเร็วกว่าเพราะอัตราการเรียนรู้มากกว่า แต่เมื่อเข้าสู่สถานะอยู่ตัว มีเพียงวิธี Entropy เท่านั้นที่ปรับอัตราการเรียนรู้ลดลง ซึ่งจะให้การประมาณค่าที่ใกล้เคียงมากขึ้น

ในรูปที่ 4.3 ถึงรูปที่ 4.6 เป็นผลการทดสอบเมื่อใช้ฟังก์ชันอ้างอิงที่มีจำนวนตัวแปรอิสระมากขึ้น เงื่อนไขที่ใช้ในการทดสอบยังคงเป็นเช่นเดียวกับในกรณีทีฟังก์ชันมี 3 ตัวแปร แต่ในกรณีที่จำนวนสัญญาณขาเข้ามากขึ้นทำให้จำนวนพารามิเตอร์มากขึ้น ซึ่งจะมีผลต่อวิธีการปรับอัตราการเรียนรู้อย่าง Idbd ซึ่งใช้จำนวนพารามิเตอร์ในการกำหนดค่าเริ่มต้น วิธีนี้จึงให้ค่าอัตราการเรียนรู้น้อยลงเมื่อพารามิเตอร์มีจำนวนมากขึ้น และมีการเปลี่ยนแปลงน้อยทำให้ค่าที่ได้ใกล้เคียงกันตลอดการทดสอบ ระบบจึงเข้าสู่สถานะอยู่ตัวช้าลง ค่า SSE จะเพิ่มขึ้นจากจำนวนเทอมที่เพิ่มขึ้น ระบบที่มีพารามิเตอร์มากกว่าและใช้อัตราการเรียนรู้เพียงค่าเดียวจะเข้าสู่สถานะอยู่ตัวช้าลง เนื่องจากการปรับพารามิเตอร์จะมีผลต่อกันมากขึ้น ส่วนวิธี Dbd และ Idbd ได้แยกอัตราการ

เรียนรู้จากกันทำให้การปรับพารามิเตอร์เป็นอิสระจากกันทำให้มีผลของจำนวนพารามิเตอร์น้อยกว่า ส่วนวิธี direction cosine ทำให้ระบบไม่ลู่เข้า เพราะว่า อัตราการเรียนรู้ที่ได้จากวิธีนี้มีค่าเป็นลบ (ไม่ได้แสดงไว้ในกราฟของอัตราการเรียนรู้) และมีค่าใกล้ 1 ทำให้ค่า SSE เพิ่มขึ้นอย่างรวดเร็วจนไม่สามารถรักษาเสถียรภาพไว้ได้

4.4 การทดสอบโดยใช้ฟังก์ชันอ้างอิงมีเทอมค่าคงตัวเปลี่ยนตามเวลา

ในหัวข้อนี้ เป็นการทดสอบและแสดงผลการทดสอบวิธีการปรับอัตราการเรียนรู้แบบต่าง ๆ เปรียบเทียบกับการใช้อัตราการเรียนรู้เป็นค่าคงตัว ในกรณีที่ฟังก์ชันอ้างอิงมีเทอมที่เป็นค่าคงตัวเปลี่ยนแปลงตามเวลา ดังนั้น เราสามารถแทนเทอมค่าคงตัวในสมการ (4.2.1) ด้วยเทอมที่ขึ้นกับเวลา ดังนี้

$$f(x) = \sum_{i=1}^{n} a_i x_i + K(t) \qquad (4.4.1)$$

และกำหนดให้มีอัตราการเปลี่ยนแปลงคงที่ได้ ดังสมการ

$$K(t) = K(0) + \frac{t}{\zeta_t} \qquad (4.4.2)$$

เมื่อ $K(0)$ เป็นเทอมค่าคงตัวเริ่มต้นและ ζ_t เป็นค่าคงตัว ถ้า ζ_t เป็นบวก $K(t)$ จะมีการเปลี่ยนแปลงเพิ่มขึ้น แต่ถ้าเป็นลบก็จะลดลง ส่วน t เป็นสัญญลักษณ์แทนเวลา ในการทดสอบจะกำหนดให้เทอมค่าคงตัวเปลี่ยนไปในแต่ละครั้งที่ข้อมูลเข้าระบบ ดังนั้น จึงกำหนดค่า ζ_t เป็นจำนวนข้อมูลเข้าระบบที่ทำให้เกิดการเปลี่ยนแปลง 1 หน่วย และใช้ t เป็นจำนวนข้อมูลที่เข้าระบบ

การทดสอบในหัวข้อนี้ ใช้การกำหนดพารามิเตอร์ให้กับวิธีการปรับอัตราการเรียนรู้ เช่นเดียวกับในกรณีที่ฟังก์ชันอ้างอิงไม่มีการเปลี่ยนแปลงตามเวลา ส่วนฟังก์ชันอ้างอิงสามารถปรับเปลี่ยนจากสมการ (4.3.1) ถึง (4.3.3) ได้โดยการเปลี่ยนเทอมค่าคงตัวที่เคยเป็น 0 ให้มีค่าเปลี่ยนตามเทอมในสมการ (4.4.2) ดังนี้

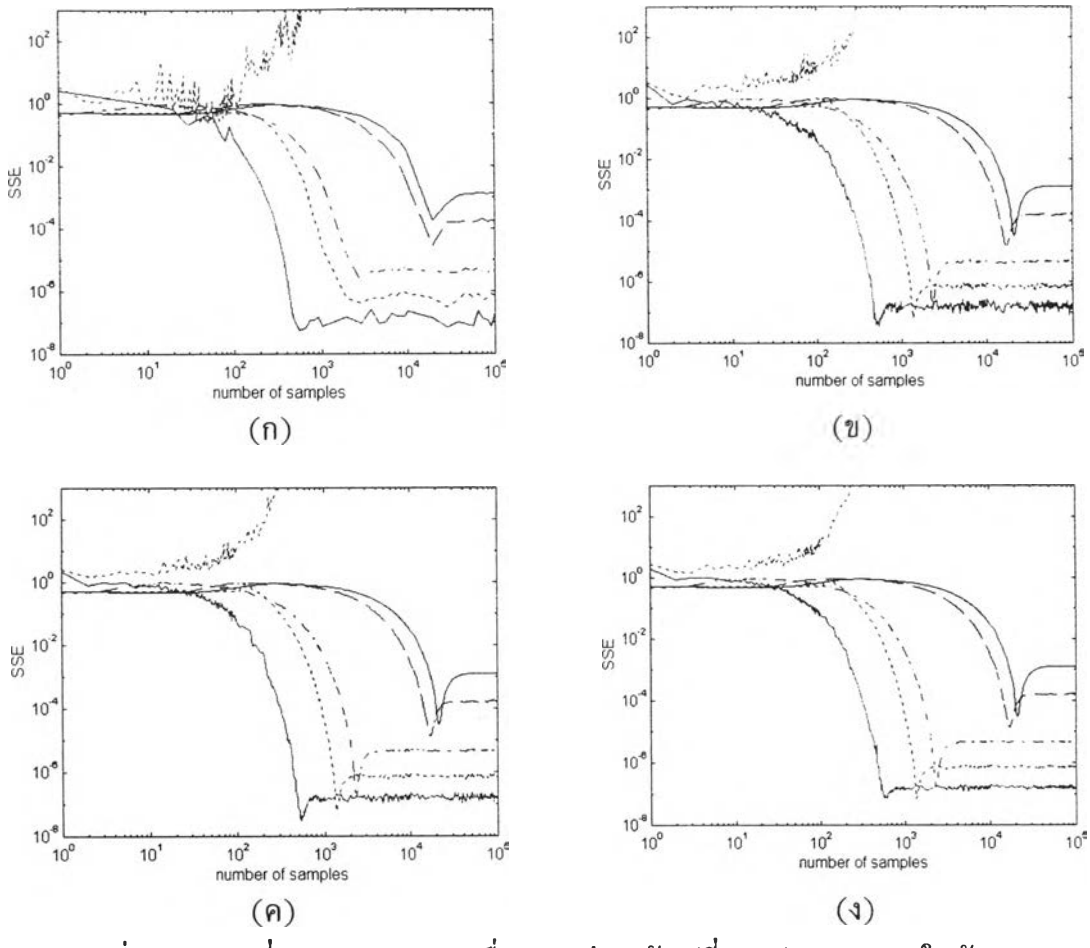
$$y_t = y + \frac{t}{\zeta_t} \qquad (4.5.3)$$

เมื่อ y_t เป็นฟังก์ชันอ้างอิงที่มีเทอมค่าคงตัวเปลี่ยนแปลงตามเวลา ซึ่งปรับเปลี่ยนรูปสมการจาก y ในสมการ (4.3.1) ถึง (4.3.3) ส่วนค่า t และ a เป็นค่าเดียวกับในสมการ (4.4.2) ข้อมูลที่ใช้ทดสอบจะเป็นค่าตัวแปรอิสระที่กระจายอย่างสม่ำเสมอในช่วง [-0.5, 0.5] และใช้ฟังก์ชันอ้างอิง y_t เพื่อหาสัญญาณอ้างอิง $d(t)$ ของระบบ

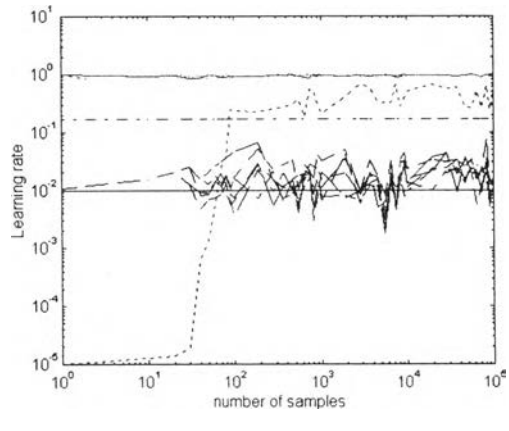
การกำหนดค่าเริ่มต้นของพารามิเตอร์ต่าง ๆ ทั้งในระบบและวิธีปรับอัตราการเรียนรู้ จะกำหนดเช่นเดียวกับการทดสอบกรณีที่ไม่มีการเปลี่ยนแปลงตามเวลา ส่วนการกำหนดอัตราการเปลี่ยนแปลง ได้ทำการทดสอบเมื่อ ζ_t มีค่าเป็น 100,000 , 50,000 และ 10,000 ข้อมูลต่อการเปลี่ยนแปลง 1 หน่วย ทำการทดสอบทั้งหมด 20 ครั้ง นำข้อมูลมาสรุปโดยการนำผลที่ได้ในแต่ละครั้งมาเฉลี่ยหลังจากทำการทดสอบไป 1, 5, 10 และ 20 ครั้ง และแสดงอัตราการเรียนรู้ของแต่ละวิธีไว้

ในการทดสอบ กำหนดให้ระบบประมาณฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 5 ตัว ผลการทดสอบได้แสดงไว้ในรูปที่ 4.7 ถึง 4.12 เพื่อให้เห็นผลจากอัตราการเปลี่ยนแปลงที่เปลี่ยนไป ในรูปที่ 4.7 และ 4.8 เป็นกราฟของ SSE และอัตราการเรียนรู้เมื่อกำหนดให้ฟังก์ชันอ้างอิงมีอัตราการเปลี่ยนแปลง 100,000 ชุดข้อมูลต่อ 1 หน่วย จากผลการทดสอบเห็นได้ว่า ในช่วงแรกกราฟเพิ่มขึ้นเล็กน้อย ซึ่งแสดงว่าระบบยังไม่สามารถติดตามการเปลี่ยนแปลงที่เกิดขึ้นได้ แต่เมื่อระบบได้รับข้อมูลมากขึ้นก็สามารถปรับตัวเข้าหาฟังก์ชันอ้างอิงทำให้ความผิดพลาดลดลง เมื่อระบบเข้าสู่สถานะอยู่ตัว ความผิดพลาดจะเพิ่มขึ้นเล็กน้อยเพราะระบบเริ่มไม่สามารถปรับตัวให้ใกล้เคียงกับฟังก์ชันได้อีกแต่ฟังก์ชันมีการเปลี่ยนแปลงตลอดเวลา เมื่อเข้าสู่สถานะอยู่ตัวแล้ว ระบบก็ไม่สามารถปรับตัวเพื่อเข้าใกล้ฟังก์ชันมากขึ้นได้อีก วิธีที่ทำให้ระบบปรับตัวได้เร็วจะทำให้ความผิดพลาดที่เพิ่มขึ้นนี้น้อยกว่า

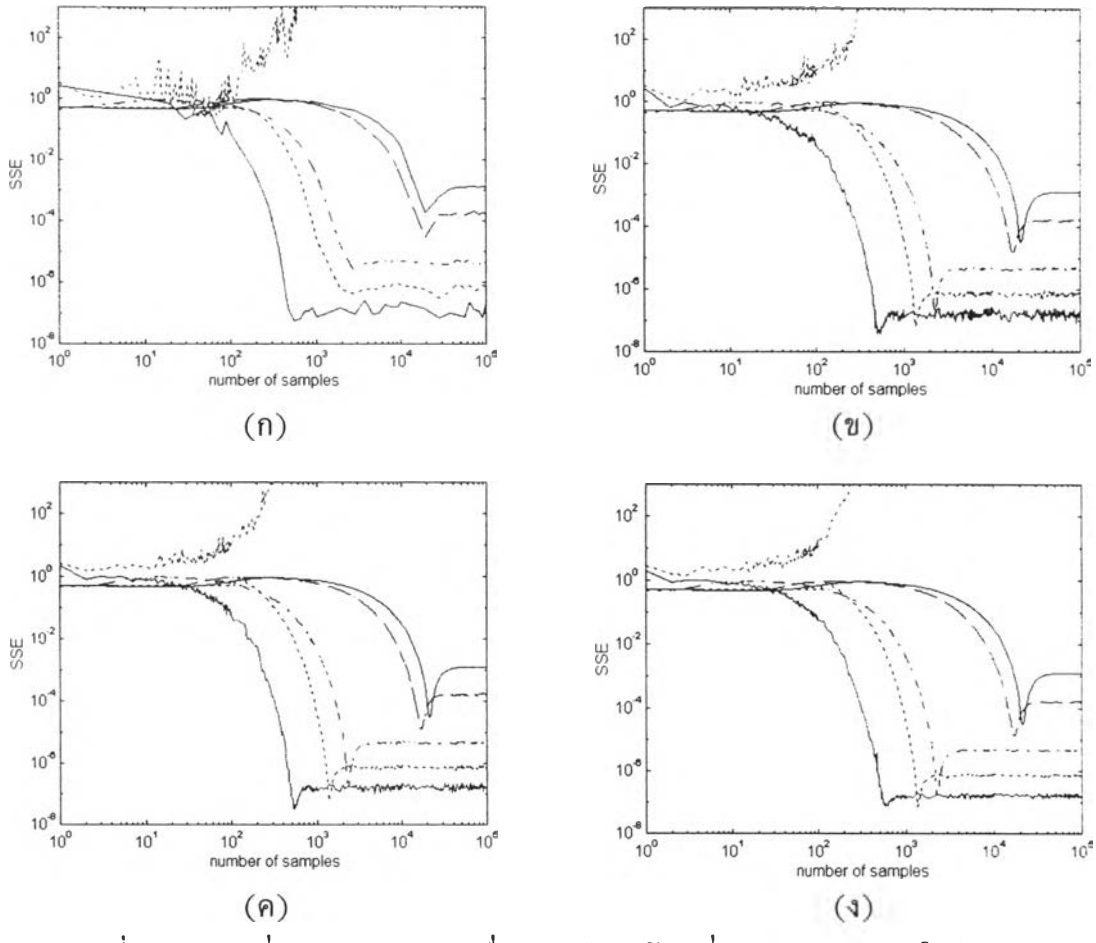
เมื่อพิจารณากราฟ SSE ร่วมกับอัตราการเรียนรู้ เห็นได้ว่า วิธี Entropy ให้อัตราการเรียนรู้ที่แตกต่างจากกรณีที่ไม่มีการเปลี่ยนแปลง โดยอัตราการเรียนรู้ไม่ลดลงเมื่อเข้าสู่สถานะอยู่ตัว เพราะว่า ฟังก์ชันอ้างอิงเคลื่อนห่างจากจุดที่ระบบปรับตัวตลอดเวลา ดังนั้น จึงจำเป็นต้องใช้อัตราการเรียนรู้ที่มีค่ามากเพื่อให้สามารถปรับตัวได้เร็ว วิธีที่สามารถให้อัตราการเรียนรู้สูง อย่าง direction cosine จึงสามารถทำให้ระบบปรับตัวเข้าใกล้ฟังก์ชันอ้างอิงได้มากกว่า



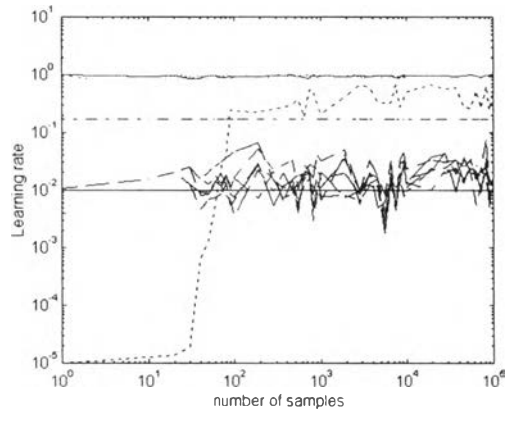
รูปที่ 4.7 ผลเฉลี่ยจากการทดสอบเมื่อเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 100,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



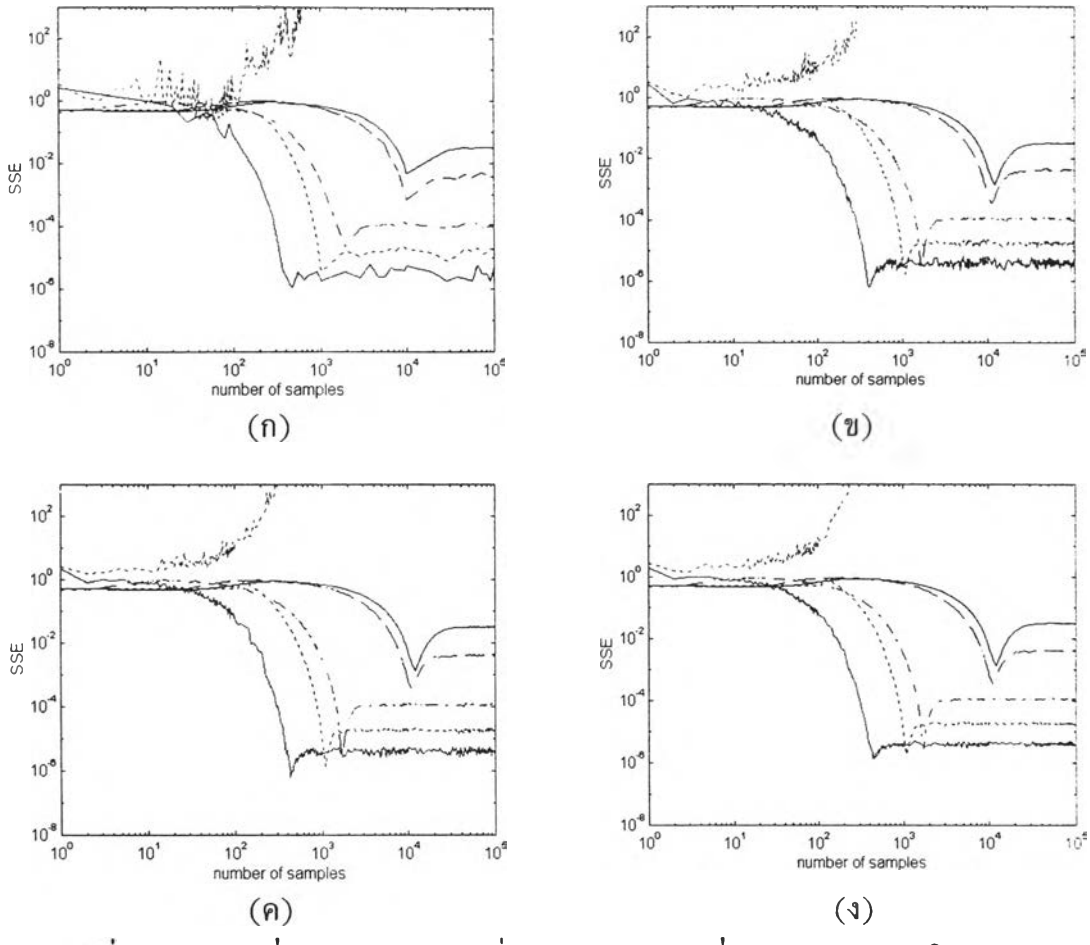
รูปที่ 4.8 อัตราการเรียนรู้เมื่อทดสอบโดยเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 100,000 ข้อมูลต่อหน่วย



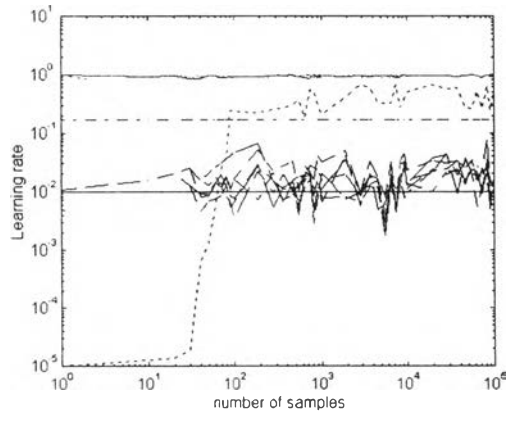
รูปที่ 4.9 ผลเฉลี่ยจากการทดสอบเมื่อเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 50,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.10 อัตราการเรียนรู้เมื่อทดสอบโดยเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 50,000 ข้อมูลต่อหน่วย



รูปที่ 4.11 ผลเฉลี่ยจากการทดสอบเมื่อเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 10,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.12 อัตราการเรียนรู้เมื่อทดสอบโดยเทอมค่าคงตัวเปลี่ยนแปลงตามเวลาในอัตรา 10,000 ข้อมูลต่อหน่วย

เมื่อกำหนดให้ฟังก์ชันอ้างอิงมีอัตราการเปลี่ยนแปลงเร็วขึ้น ผลการทดสอบได้แสดงไว้ในรูปที่ 4.9 และ 4.10 ซึ่งเป็นผลจากการใช้อัตราการเปลี่ยนแปลงที่ 50,000 ข้อมูลต่อหน่วย และรูปที่ 4.11 และ 4.12 เป็นผลเมื่อใช้อัตราการเปลี่ยนแปลงที่ 10,000 ข้อมูลต่อหน่วย จากผลการทดสอบ เมื่อการเปลี่ยนแปลงเร็วขึ้น ความผิดพลาดจะเพิ่มขึ้นทั้งตอนเริ่มปรับตัวและเมื่ออยู่ในสถานะอยู่ตัว เพราะระบบไม่สามารถติดตามการเปลี่ยนแปลงที่เร็วขึ้นได้ทัน ซึ่งสังเกตได้จากอัตราการเรียนรู้ในขณะปรับตัวที่ได้จากแต่ละวิธี ยังคงให้ค่าที่ใกล้เคียงกับในกรณีที่มีการเปลี่ยนแปลงช้ากว่า ส่วนวิธี direction cosine ให้อัตราการเรียนรู้เป็นค่าลบทำให้ค่า SSE เพิ่มขึ้นและทำให้ระบบไม่ลู่เข้าเช่นเดียวกับในกรณีที่ฟังก์ชันไม่มีการเปลี่ยนแปลงตามเวลา

4.5 การทดสอบโดยใช้ฟังก์ชันอ้างอิงมีสัมประสิทธิ์เปลี่ยนตามเวลา

ในหัวข้อที่ 4.4 เรากำหนดให้เทอมค่าคงตัวของฟังก์ชันอ้างอิงมีค่าเปลี่ยนแปลงตามเวลา ดังนั้น ในหัวข้อนี้ เราจะทดสอบให้ระบบปรับตัวเพื่อประมาณฟังก์ชันอ้างอิงที่มีสัมประสิทธิ์ของตัวแปรอิสระมีค่าเปลี่ยนตามเวลา ฟังก์ชันอ้างอิงในหัวข้อนี้จะสามารถเขียนในรูปทั่วไปได้เป็น

$$f(x) = \sum_{i=1}^{n} a_i(t)x_i + K \qquad (4.5.1)$$

และกำหนดให้ค่าสัมประสิทธิ์ $a_i(t);\ 1 \le i \le N$ เปลี่ยนตามเวลาโดยให้มีอัตราการเปลี่ยนแปลงคงที่ ดังนั้นค่าสัมประสิทธิ์ในแต่ละช่วงเวลาจะเป็นไปตามสมการ

$$a_i(t) = a_i(0) + \frac{t}{\zeta} \qquad (4.5.2)$$

เมื่อ $a_i(0)$ เป็นค่าสัมประสิทธิ์เริ่มต้นและ ζ เป็นอัตราการเปลี่ยนแปลงเช่นเดียวกับในหัวข้อที่ 4.4 ส่วน t เป็นสัญญลักษณ์แทนเวลา โดยกำหนดให้ค่าสัมประสิทธิ์เปลี่ยนไปในแต่ละครั้งที่ข้อมูลเข้าระบบ ดังนั้น ค่า ζ และ t จะสามารถกำหนดได้เช่นเดียวกับในหัวข้อที่ 4.4 การกำหนดค่าพารามิเตอร์สำหรับแต่ละวิธียังคงเป็นเช่นเดียวกับในกรณีที่ฟังก์ชันอ้างอิงไม่มีการเปลี่ยนแปลง

ฟังก์ชันอ้างอิงสามารถปรับเปลี่ยนจากสมการ (4.3.1) ถึง (4.3.3) ได้โดยการเปลี่ยนค่าสัมประสิทธิ์ที่มีค่าเป็น 1 ให้มีค่าเปลี่ยนตามสมการ (4.5.2) ซึ่งค่า $a_i(0) = 1$ ดังนั้น ฟังก์ชันอ้างอิงในกรณีที่มีตัวแปรอิสระ 5 ตัว สามารถเขียนได้เป็น

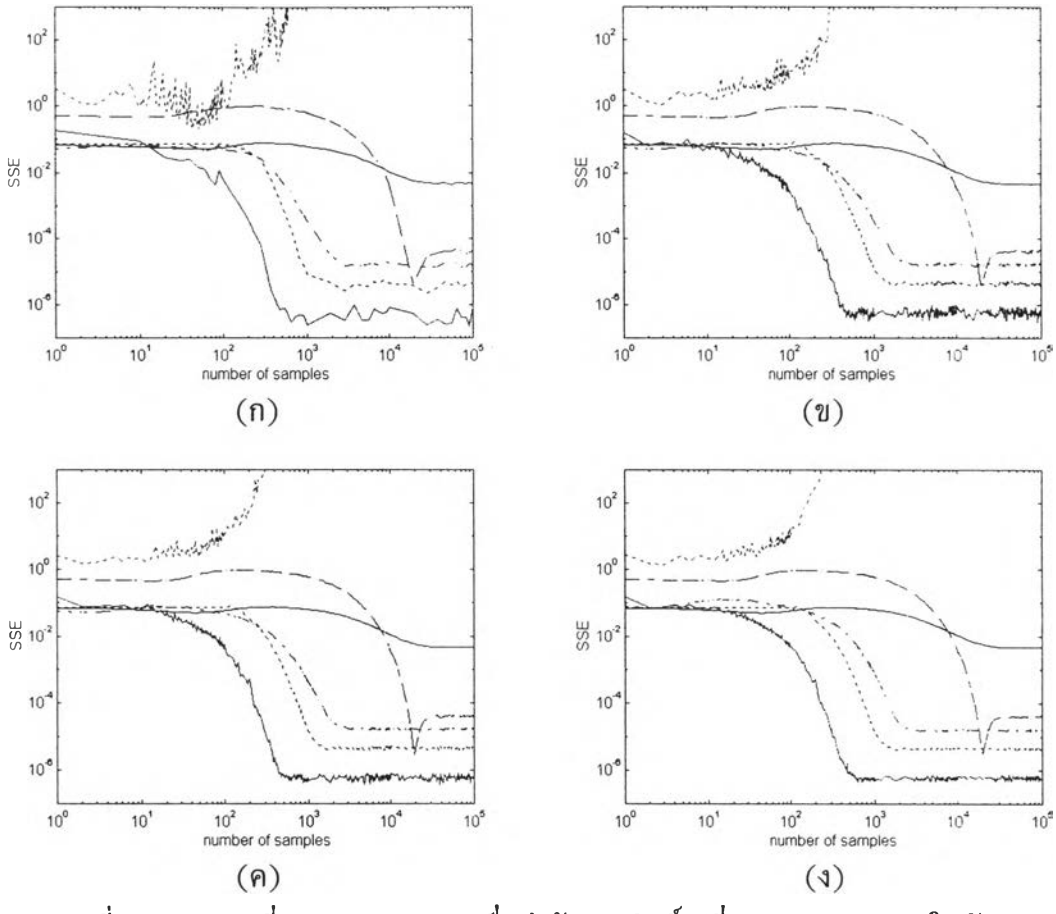
$$y_r = a_1(t)x_1 + a_2(t)x_2 + a_3(t)x_3 + a_4(t)x_4 + a_5(t)x_5 \qquad (4.5.3)$$

ในการทดสอบ ข้อมูลที่ใช้เป็นค่าตัวแปรอิสระที่กระจายอย่างสม่ำเสมอในช่วง [-0.5, 0.5] และใช้ฟังก์ชันอ้างอิง y_r สำหรับหาสัญญาณอ้างอิง $d(t)$ ของระบบ เช่นเดียวกัน การกำหนดค่าเริ่มต้นของพารามิเตอร์ต่าง ๆ ทั้งในระบบและวิธีปรับอัตราการเรียนรู้ จะกำหนดเช่นเดียวกับการทดสอบกรณีที่ไม่มีการเปลี่ยนแปลงตามเวลา ส่วนอัตราการเปลี่ยนแปลงที่ใช้ ได้เลือกใช้ค่าเดียวกับในหัวข้อที่ 4.4 คือ ζ มีค่าเป็น 100,000 , 50,000 และ 10,000 ข้อมูลต่อการเปลี่ยนแปลง 1 หน่วย การทดสอบจะทำทั้งหมด 20 ครั้งและนำมาเฉลี่ยเมื่อทำการทดสอบไป 1, 5, 10 และ 20 ครั้ง และแสดงอัตราการเรียนรู้ของแต่ละวิธีไว้

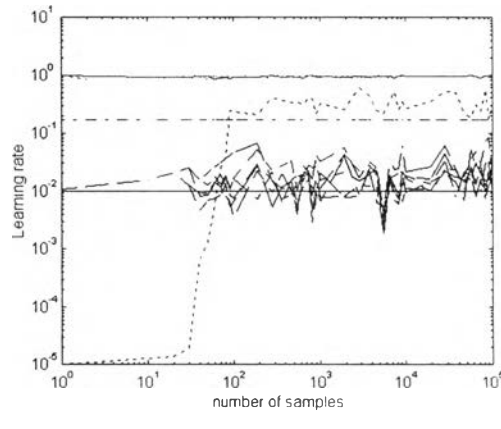
การทดสอบให้ค่าสัมประสิทธิ์ของฟังก์ชันอ้างอิงเปลี่ยนแปลงตามเวลานี้ จะกำหนดให้สัมประสิทธิ์ทุกตัวเปลี่ยนแปลงไปพร้อม ๆ กันตามอัตราที่กำหนดไว้ ผลการทดสอบได้แสดงไว้ในรูปที่ 4.13 ถึง 4.18 ซึ่งแสดงให้เห็นถึงกรณีที่ระบบไม่สามารถติดตามการเปลี่ยนแปลงของฟังก์ชันอ้างอิงได้ทัน เมื่อใช้อัตราการเรียนรู้เป็นค่าคงตัวที่ 0.01 ส่วนการใช้วิธีปรับอัตราการเรียนรู้แบบต่าง ๆ ยังคงสามารถติดตามการเปลี่ยนแปลงได้ แต่มีประสิทธิภาพต่างกัน ในการทดสอบนี้วิธีที่ใช้อัตราการเรียนรู้มากอย่าง modified direction cosine ยังคงสามารถทำให้ระบบปรับตัวเข้าหาฟังก์ชันอ้างอิงได้ดีที่สุด ถัดมาคือวิธี Entropy ซึ่งสามารถปรับอัตราการเรียนรู้ให้มีการเปลี่ยนแปลงตามการใช้งานได้ดี โดยค่าอัตราการเรียนรู้จะไม่ลดลงหลังจากเข้าสู่สถานะอยู่ตัว เช่นเดียวกับกรณีที่เทอมค่าคงตัวมีการเปลี่ยนแปลง

เมื่อพิจารณากราฟผลรวมความผิดพลาดกำลังสอง เห็นได้ว่า กราฟของวิธีการปรับอัตราการเรียนรู้สามารถลดลงและเข้าสู่สถานะอยู่ตัวได้เร็ว โดยที่วิธีที่ให้อัตราการเรียนรู้มากกว่าจะสามารถทำให้ระบบปรับตัวได้เร็วกว่า แต่ว่าการให้อัตราการเรียนรู้ที่มีค่าสูงจะทำให้เกิดการแกว่งเมื่อเข้าสู่สถานะอยู่ตัวได้ ดังกราฟของวิธี direction cosine ซึ่งจะมีการแกว่งในสถานะอยู่ตัวมากกว่าวิธีอื่น ๆ ส่วนวิธีที่ให้ผลแตกต่างวิธีอื่น ๆ อีก ได้แก่ วิธี Dbd ซึ่งให้กราฟที่มีการเพิ่มขึ้นในตอนเริ่มและก่อนเข้าสู่สถานะอยู่ตัว เพราะว่า วิธีนี้ให้อัตราการเรียนรู้ที่มีการแกว่งไปมา ทำให้การติดตามเกิดความไม่แน่นอนในการปรับตัวของระบบ ซึ่งในกรณีที่การเปลี่ยนแปลงเร็ว จะยิ่งเห็นผลได้ชัดเจนขึ้น

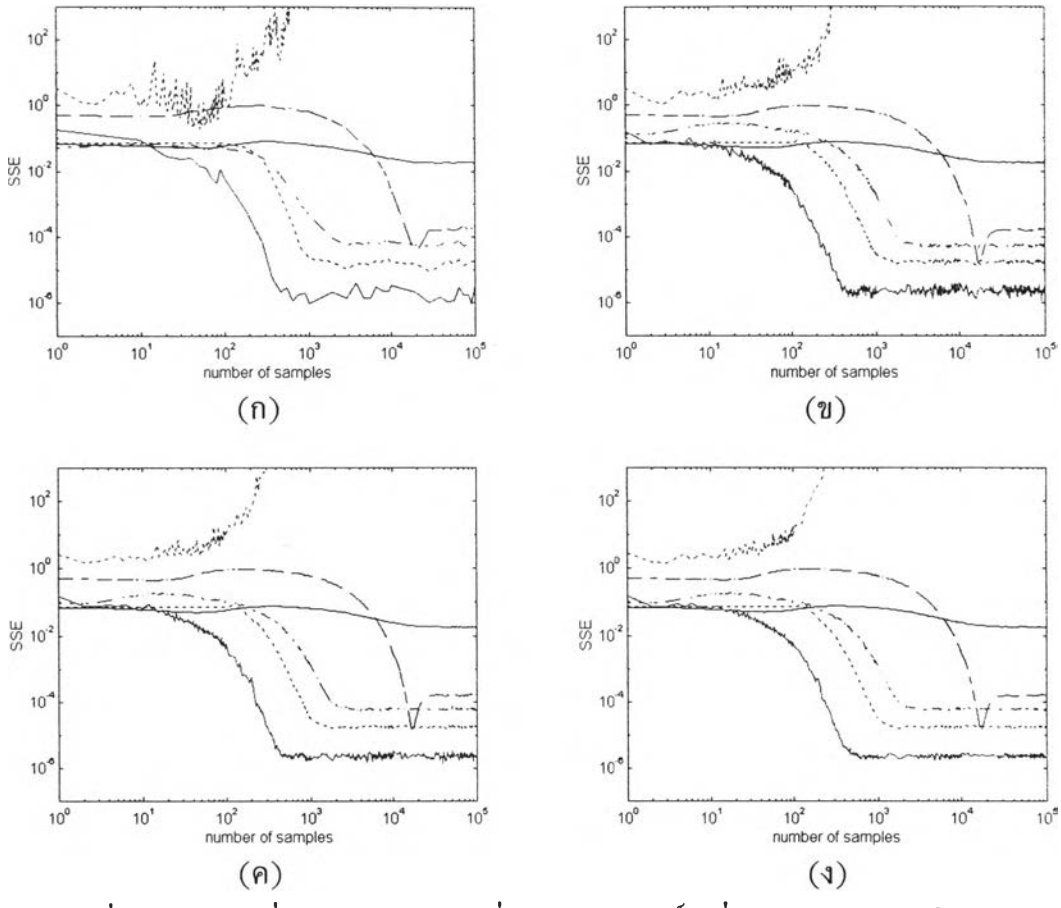
เส้นกราฟที่เพิ่มขึ้นแสดงผลจากการใช้วิธี direction cosine เพื่อปรับอัตราการเรียนรู้ อัตราการเรียนรู้ที่ได้เป็นลบและมีค่าใกล้ 1 ทำให้ค่า SSE เพิ่มขึ้นอย่างรวดเร็ว และไม่ลดลงอีก เพราะว่าวิธีนี้ให้ค่าที่เป็นลบสลับกับค่าที่เป็นบวกตลอดการปรับตัวของระบบ



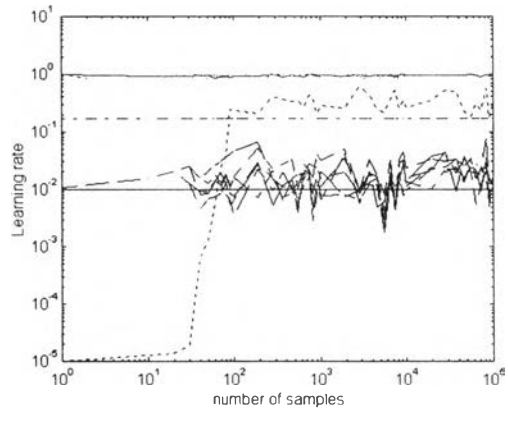
รูปที่ 4.13 ผลเฉลี่ยจากการทดสอบเมื่อค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 100,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



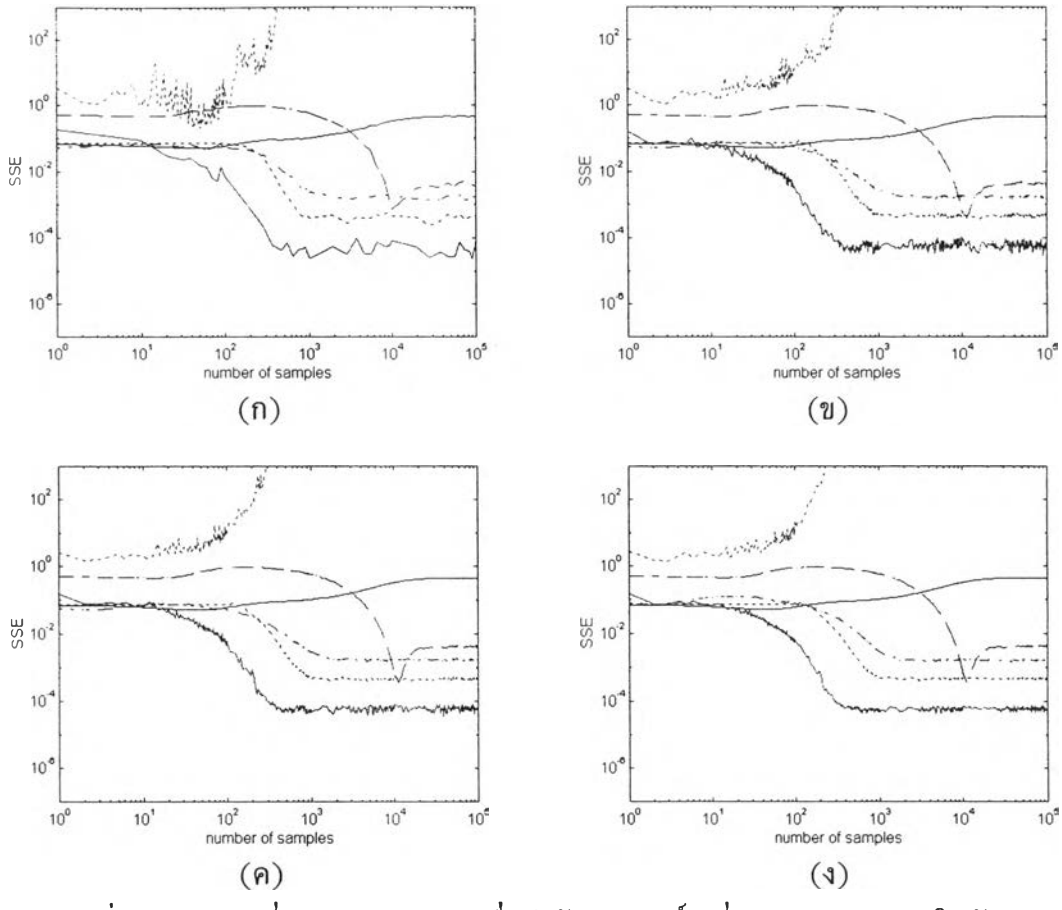
รูปที่ 4.14 อัตราการเรียนรู้เมื่อทดสอบโดยค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 100,000 ข้อมูลต่อหน่วย



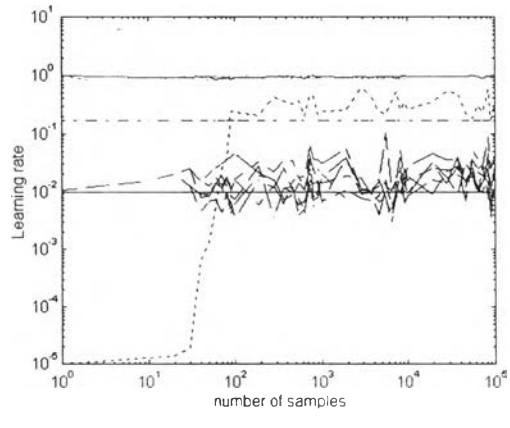
รูปที่ 4.15 ผลเฉลี่ยจากการทดสอบเมื่อค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 50,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.16 อัตราการเรียนรู้เมื่อทดสอบโดยค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 50,000 ข้อมูลต่อหน่วย



รูปที่ 4.17 ผลเฉลี่ยจากการทดสอบเมื่อค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 10,000 ข้อมูลต่อหน่วย (ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



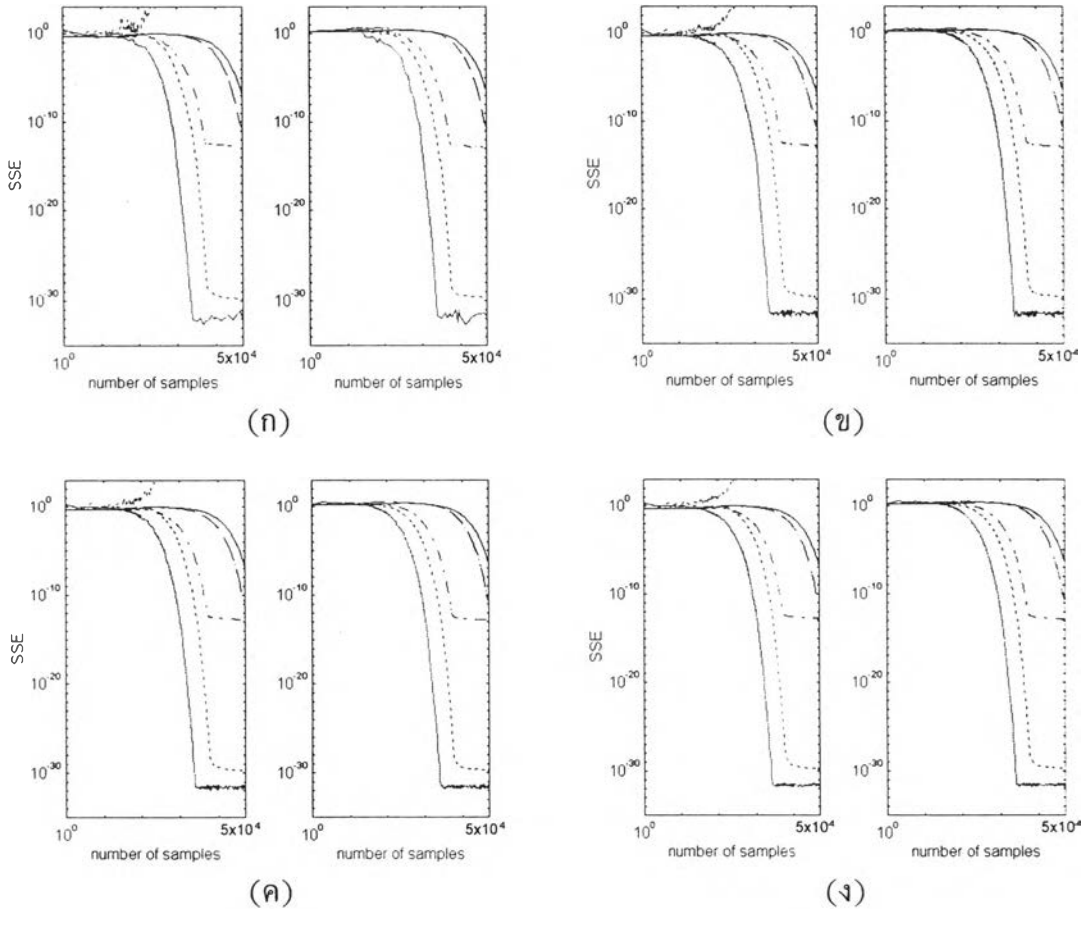
รูปที่ 4.18 อัตราการเรียนรู้เมื่อทดสอบโดยค่าสัมประสิทธิ์เปลี่ยนแปลงตามเวลาในอัตรา 10,000 ข้อมูลต่อหน่วย

4.6 การทดสอบกับฟังก์ชันอ้างอิงที่มีการเปลี่ยนแปลงแบบทันทีทันใด

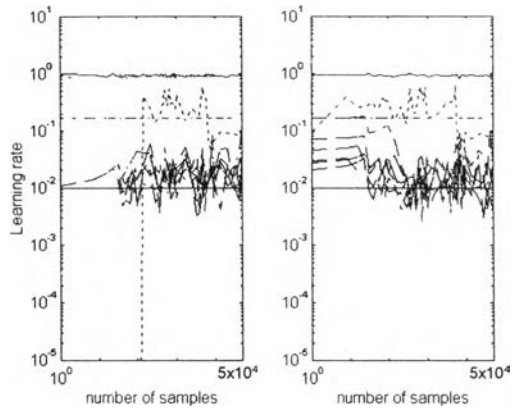
ในหัวข้อนี้ จะทำการทดสอบกับฟังก์ชันอ้างอิงที่มีการเปลี่ยนแปลงแบบทันทีทันใด หลังจากที่ระบบเข้าสู่สถานะอยู่ตัวแล้ว (หลังจากระบบได้รับข้อมูล 50,000 ชุด) การเปลี่ยนแปลงนี้จะกำหนดโดยให้ฟังก์ชันอ้างอิง มีค่าเป็นลบจากค่าที่ควรเป็น ซึ่งแสดงได้ดังสมการ

$$f(x) = -\left(\sum_{i=1}^{n} a_i x_i + K\right) \tag{4.6.1}$$

การกำหนดพารามิเตอร์ให้กับวิธีการปรับอัตราการเรียนรู้และค่าเริ่มต้นต่าง ๆ ในการทดสอบยังคงเป็นเช่นเดียวกับในกรณีที่ฟังก์ชันอ้างอิงไม่มีการเปลี่ยนแปลงตามเวลา



รูปที่ 4.19 ผลเฉลี่ยจากการทดสอบกับฟังก์ชันอ้างอิงที่มีการเปลี่ยนแปลงแบบทันทีทันใด
(ก) 1 ครั้ง (ข) 5 ครั้ง (ค) 10 ครั้ง (ง) 20 ครั้ง



รูปที่ 4.20 อัตราการเรียนรู้เมื่อทดสอบกับฟังก์ชันอ้างอิงที่มีการเปลี่ยนแปลงแบบทันทีทันใด

ผลการทดสอบจะแสดงเป็นกราฟ SSE และอัตราการเรียนรู้เหมือนที่ผ่านมา แต่เนื่องจากต้องการแสดงให้เห็นถึงการเปลี่ยนแปลงของกราฟ หลังจากที่ได้ปรับเปลี่ยนฟังก์ชันอ้างอิงแล้ว จึงได้แบ่งกราฟ SSE เป็น 2 ส่วน ส่วนทางด้านซ้ายจะแสดงผลการปรับตัวเมื่อระบบได้ข้อมูล 50,000 ชุด และทางด้านขวาแสดงผลหลังจากที่ฟังก์ชันมีการเปลี่ยนแปลงแล้วและระบบได้รับข้อมูลอีก 50,000 ชุด

การทดสอบได้ใช้ฟังก์ชันอ้างอิงที่มีตัวแปรอิสระ 5 ตัวเป็นตัวอย่าง และผลการทดสอบได้แสดงในรูปที่ 4.19 และ 4.20 เห็นได้ว่า ระบบสามารถกลับเข้าสู่สถานะอยู่ตัวได้อีกครั้งหลังจากมีการเปลี่ยนแปลงฟังก์ชันอ้างอิง ระบบที่ใช้วิธีการปรับอัตราการเรียนรู้แบบต่าง ๆ สามารถปรับตัวให้กลับเข้าสู่สถานะอยู่ตัวได้เร็วกว่าระบบที่ใช้อัตราการเรียนรู้เป็นค่าคงตัว วิธี modified direction cosine สามารถทำให้ระบบปรับได้เร็วกว่าวิธีอื่น ๆ จากอัตราการเรียนรู้ที่มากกว่า และยังให้การประมาณที่ใกล้เคียงฟังก์ชันอ้างอิงมากกว่า จากอัตราการเรียนรู้ที่ลดลงเมื่อเข้าสู่สถานะอยู่ตัว

เมื่อมีการเปลี่ยนแปลงเกิดขึ้น อัตราการเรียนรู้ของวิธี Dbd และ Entropy ถูกปรับเพิ่มขึ้นเพื่อให้ระบบปรับตัวให้ใกล้เคียงกับฟังก์ชันที่เปลี่ยนไปได้เร็วมากขึ้นและลดอัตราการเรียนรู้เมื่อระบบเข้าสู่สถานะอยู่ตัวอีกครั้ง ในขณะที่อัตราการเรียนรู้ของวิธี Idbd มีการเปลี่ยนแปลงน้อยมาก

4.7 สรุป

ในบทนี้ แสดงการทดสอบการปรับอัตราการเรียนรู้ในข่ายงานระบบประสาทแบบชั้นเดียวกับการประมาณฟังก์ชันเชิงเส้น ซึ่งเป็นปัญหาที่ง่ายและชัดเจน ผลการทดสอบได้แสดงว่า วิธีการ

ปรับที่ให้อัตราการเรียนรู้ได้สูงกว่า จะสามารถทำให้ระบบปรับตัวได้เร็วกว่า และเนื่องจากปัญหาที่ใช้เป็นปัญหาที่ง่าย โดยข่ายงานระบบประสาทที่ใช้สามารถประมาณฟังก์ชันอ้างอิงที่เป็นเชิงเส้นได้อย่างแม่นยำ จึงทำให้วิธีการปรับที่ให้อัตราการเรียนรู้ที่สูงกว่าสามารถให้ค่าผลรวมความผิดพลาดกำลังสองได้น้อยกว่าด้วย

วิธีที่ให้ผลได้ดีที่สุดสำหรับปัญหาในบทนี้ คือวิธี modified direction cosine ซึ่งได้รับปรับปรุงให้มีอัตราการเรียนรู้เป็นบวกเสมอ เพราะว่า สามารถให้อัตราการเรียนรู้ที่มีค่าใกล้ 1 ตลอดการปรับตัวของระบบ ส่วนวิธีที่สามารถให้ประสิทธิภาพได้ใกล้เคียงกัน คือวิธี Entropy ซึ่งสามารถปรับอัตราการเรียนรู้ให้เพิ่มขึ้นเพื่อให้ระบบปรับตัวได้เร็ว และลดอัตราการเรียนรู้ลงในสถานะอยู่ตัวเพื่อให้ระบบสามารถปรับตัวให้ใกล้เคียงมากขึ้นได้ หรือในกรณีที่มีฟังก์ชันอ้างอิงมีการเปลี่ยนแปลงตามเวลา วิธีนี้ก็ปรับอัตราการเรียนรู้ให้มีค่าสูงตลอดเพื่อให้ระบบติดตามการเปลี่ยนแปลงที่เกิดขึ้นได้ เมื่อเปรียบเทียบกับวิธีอื่น ๆ ซึ่งไม่มีการเปลี่ยนแปลงของอัตราการเรียนรู้ให้เห็น วิธีนี้จึงน่าจะสามารถปรับปรุงประสิทธิภาพของระบบได้มาก แต่เพียงไม่สามารถให้อัตราการเรียนรู้ได้มากเท่ากับวิธี modified direction cosine เท่านั้น ส่วนวิธี direction cosine ที่ยังไม่มีการปรับปรุงให้ค่าอัตราการเรียนรู้ที่เป็นลบสลับกับค่าบวก ทำให้ระบบที่ใช้วิธีนี้ไม่ลู่เข้า

การทดสอบในบทนี้ วิธี Dbd ให้อัตราการเรียนรู้ที่แกว่งไปมา ทำให้ระบบไม่สามารถปรับตัวได้อย่างแน่นอน จึงทำให้มีการเปลี่ยนแปลงของกราฟผลรวมความผิดพลาดกำลังสอง โดยเฉพาะในกรณีที่ฟังก์ชันอ้างอิงเปลี่ยนแปลงตามเวลา วิธีนี้จะทำให้ผลรวมความผิดพลาดกำลังสองเพิ่มขึ้นก่อนเข้าสู่สถานะอยู่ตัวมากกว่าวิธีอื่น ๆ ซึ่งการแกว่งของอัตราการเรียนรู้ที่เกิดขึ้นนี้ อาจเกิดจากการเลือกชุดค่าพารามิเตอร์ที่ยังไม่เหมาะสมกับการใช้งานได้เช่นกัน วิธี Idbd ให้อัตราการเรียนรู้ที่เกือบคงตัวตลอดการปรับตัวของระบบ ซึ่งจะมีผลเหมือนการใช้อัตราการเรียนรู้เป็นค่าคงตัว แต่ว่าวิธีนี้มีการเลือกค่าอัตราการเรียนรู้อย่างอัตโนมัติ ซึ่งเหมาะสมกับการใช้งานมากกว่าเลือกใช้ค่าคงตัวที่ 0.01 และปัญหาที่ใช้ในการทดสอบไม่มีการเปลี่ยนแปลงมากนัก จึงทำให้อัตราการเรียนรู้ที่ได้จากวิธีนี้ใช้งานได้ดีพอสมควร

ผลการทดสอบในบทนี้ สามารถเป็นแนวทางสำหรับศึกษาลักษณะการปรับอัตราการเรียนรู้ของแต่ละวิธีได้อย่างดี เช่น วิธี Idbd จะมีการเปลี่ยนแปลงน้อย วิธี Dbd อัตราการเรียนรู้จะขึ้นกับการกำหนดค่าเริ่มต้น และการเลือกค่าพารามิเตอร์ของขั้นตอน วิธี Entropy ให้การเปลี่ยนแปลงของอัตราการเรียนรู้ได้ดีและเหมาะสมกับการใช้งาน ส่วนวิธี direction cosine จะให้อัตราการเรียนรู้ที่มีค่าสูง ซึ่งข้อมูลเหล่านี้เกิดจากการทดสอบใช้งานจริงกับปัญหาเบื้องต้น ซึ่งทำให้เห็นการทำงานได้ชัดเจนกว่า เมื่อนำไปประกอบกับแนวทางการสร้างขั้นตอน สามารถเป็นข้อมูลสำหรับการเลือกใช้วิธีปรับอัตราการเรียนรู้ให้เหมาะสมกับการใช้งานได้เป็นอย่างดี