บทที่ 3
การวิเคราะห์โครงสร้างของดีดีเอจีและเอดีเอจี

ในบทนี้จะกล่าวถึงปัญหาของดีดีเอจี 2 ประเด็นได้แก่ (1) การที่ผลการจำแนกขึ้นกับลำดับของกลุ่มในลิสต์ และ (2) ระดับความลึกของเส้นทางจำแนกของดีดีเอจี จากปัญหาดังกล่าว จึงเสนอสถาปัตยกรรมใหม่สำหรับการจำแนกแบบหลายกลุ่มที่เรียกว่าเอดีเอจีและแนะนำวิธีการนำขั้นตอนวิธีเอดีเอจีไปใช้ และในส่วนสุดท้ายจะเสนอแบบจำลองสำหรับหาค่าคาดหวังของความถูกต้องของขั้นตอนวิธีทั้งสองเพื่อการเปรียบเทียบ

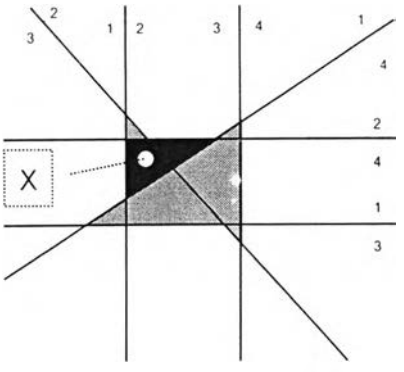
3.1 ปัญหาอันเกิดจากโครงสร้างของดีดีเอจี

3.1.2 ผลการจำแนกขึ้นกับของลำดับของกลุ่มในลิสต์

ดีดีเอจีได้รับการออกแบบอย่างดี จึงเหนือกว่าขั้นตอนการจำแนกแบบหลายกลุ่มมาตรฐานในด้านความเร็ว อย่างไรก็ตามการขึ้นกับลำดับของโนดของดีดีเอจี ส่งผลกระทบต่อความถูกต้องและความเชื่อถือได้ของผลการจำแนก

ดีดีเอจีให้ผลลัพธ์ที่มีการแจกแจงความน่าจะเป็นไม่ยูนิฟอร์ม อันเนื่องมาจากการขึ้นอยู่กับลำดับของกลุ่มในลิสต์ ทำให้ความแปรปรวนของผลการจำแนกเพิ่มขึ้น นอกจากนี้ถ้ากลุ่มที่ถูกต้องถูกเปรียบเทียบตั้งแต่ระดับบน จะทำให้มีโอกาสถูกกำจัดออกไปอย่างผิดพลาดมากกว่าในกรณีที่กลุ่มที่ถูกต้องเริ่มถูกเปรียบในระดับล่าง

ลองพิจารณาปัญหาแรกโดยใช้ตัวอย่างในรูปที่ 7 ซึ่งเป็นปัญหาการจำแนกกลุ่ม 4 กลุ่มในระนาบ 2 มิติ โดยที่เส้นแต่ละเส้นเป็นเส้นแบ่งระหว่างกลุ่มสองกลุ่ม ตัวเลขกำกับเส้นแสดงชื่อกลุ่มในฝั่งแต่ละฝั่ง เช่นเส้น 1 2 หมายความว่า จุดใดๆที่ทางซ้ายของเส้นไม่น่าจะเป็นกลุ่มที่ 2 และจุดใดๆที่อยู่ทางขวาของเส้น 1 2 ไม่น่าจะเป็นกลุ่ม 1



รูปที่ 7 ตัวอย่างแสดงผลของการจำแนกขึ้นกับลำดับของกลุ่ม

ปัญหาที่เกิดขึ้นคือ จุดที่อยู่ในบริเวณสีดำหรือสีเทาจะถูกจำแนกได้หลายแบบขึ้นอยู่กับลำดับของกลุ่มในโนด (หรือลำดับของสมาชิกในลิสต์ที่ใช้ในการจำแนก) พิจารณาตัวอย่างที่จุด x ในบริเวณสีดำ ซึ่งจะถูกจำแนกได้หลายแบบ ขึ้นอยู่กับลิสต์ที่ใช้

ในกรณีที่ใช้ลิสต์ 1-2-3-4 ตอนแรกระนาบ 1vs4 จะกำจัดกลุ่ม 4 ออกไปเนื่องจากจุด x ไม่ได้อยู่ฝั่งของกลุ่ม 4 ลิสต์จะกลายเป็น 1-2-3 ถัดมา ระนาบ 1vs3 จะเอากลุ่ม 3 ออกจากลิสต์ สุดท้ายระนาบ 1vs2 จะตัดกลุ่ม 1 ออกไป ทำให้จุด x ถูกจำแนกเป็นกลุ่ม 2

ในอีกกรณีของลิสต์ 2-1-3-4 ตอนแรกระนาบ 2vs4 จะกำจัดกลุ่ม 2 ออกไป ลิสต์จะกลายเป็น 1-3-4 ถัดมา ระนาบ 1vs4 จะเอากลุ่ม 4 ออกจากลิสต์ สุดท้ายระนาบ 1vs3 จะตัดกลุ่ม 3 ออกไป ทำให้จุด x ถูกจำแนกเป็นกลุ่ม 1 ซึ่งต่างจากกรณีแรก

ตัวอย่างนี้แสดงให้เห็นว่าการเปลี่ยนลำดับของสมาชิกในลิสต์ที่ใช้ในการจำแนกมีผลต่อผลการจำแนก ซึ่งจะกระทบต่อความถูกต้องในการจำแนกในที่สุด

ตารางที่ 3 แสดงให้เห็นว่า ในกรณีของดีดีเอจีผลการจำแนกขึ้นกับลำดับของสมาชิกของลิสต์มาก เช่นกรณีปัญหาแบบ 4 กลุ่ม จะมีรูปแบบลำดับของสมาชิกทั้งหมด 24 แบบ โดยมีแบบที่แตกต่างกันอย่างแท้จริงจำนวน 12 แบบ พบว่าเมื่อเปลี่ยนลำดับของสมาชิกไป ผลการตัดสินใจของตัวจำแนกแบบดีดีเอจีจะเปลี่ยนตามไปด้วย ทำให้ผลการจำแนกกระจายไปในกลุ่มทั้งสี่ แทนที่จะเลือกตอบกลุ่มใดกลุ่มหนึ่ง ตัวอย่างนี้แสดงถึงปัญหาประเด็นแรกของการขึ้นอยู่กับลำดับของโนดของดีดีเอจี ซึ่งทำให้ผลการจำแนกมีความน่าเชื่อถือลดลง

ตารางที่ 3 โอกาสที่จะจำแนกจุด 'x' ในรูปที่ 7 เป็นกลุ่มต่างๆ โดยลิสต์ที่มีสมาชิก 4 ตัว

ผลการจำแนก	ความน่าจะเป็น
กลุ่ม 1	41.67%
กลุ่ม 2	25.00%
กลุ่ม 3	25.00%
กลุ่ม 4	8.33%

3.1.2 ระดับความลึกของเส้นทางจำแนกของดีดีเอจี

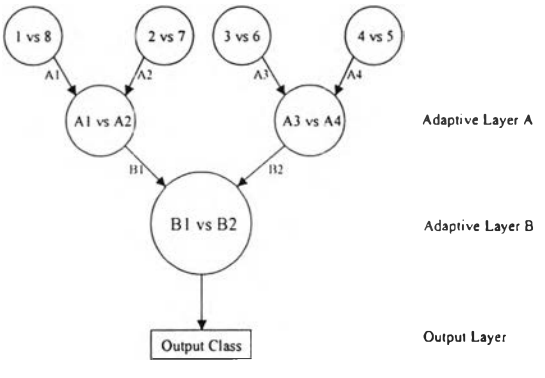
เส้นทางการจำแนกของดีดีเอจีมีความลึกโดยเฉลี่ยค่อนข้างมาก ทำให้ความผิดพลาดสะสมสูงขึ้น และส่งผลให้กลุ่มที่ถูกต้องอาจถูกกำจัดออกอย่างผิดพลาดได้ อันจะทำให้ความผิดพลาดรวมสูงและทำให้ความถูกต้องลดต่ำลง ความลึกของดีดีเอจีเป็น N-1 เมื่อ N เป็นจำนวนกลุ่มของปัญหา ซึ่งโดยเฉลี่ยแล้วจำนวนการเปรียบเทียบจะขึ้นกับ N พิจารณาตัวอย่าง

สำหรับกรณีปัญหา 20 กลุ่ม ถ้ากลุ่มที่ถูกต้องถูกเปรียบเทียบครั้งแรกที่โนดราก กลุ่มนี้จะต้องถูกเปรียบเทียบทั้งหมด 19 ครั้ง ก่อนที่จะได้รับการจำแนกอย่างถูกต้อง ถ้ากำหนดให้โอกาสผิดพลาดเป็น p ในการจำแนกแต่ละครั้ง จะทำให้ความผิดพลาดสะสมเป็น $1-(1-p)^{19}$ เช่นถ้า p = 1% จะก่อให้เกิดความผิดพลาดรวมถึง 17.38% ปัญหานี้จะรุนแรงขึ้นเมื่อจำนวนกลุ่มเพิ่มขึ้น ปัญหาทั้งสองประเด็นจูงใจให้คิดค้นวิธีปรับโครงสร้างของดีดีเอจีเพื่อพัฒนาขั้นตอนวิธีที่มีประสิทธิภาพดีขึ้นกว่าเดิม

3.2 สถาปัตยกรรมของเอดีเอจี

ส่วนนี้จะแนะนำสถาปัตยกรรมใหม่ที่ลดข้อเสียของโครงสร้างดีดีเอจีลง โครงสร้างใหม่ที่เรียกว่ากราฟไม่วงแบบมีทิศทางที่ปรับได้ (Adaptive Directed Acyclic Graph) หรือเอดีเอจี สามารถลดการขึ้นกับลำดับของสมาชิกในลิสต์และลดความลึกของเส้นทางการจำแนกลง

เอดีเอจีเป็นกราฟไม่วงแบบมีทิศทางที่มีโครงสร้างคล้ายสามเหลี่ยมคว่ำ สำหรับปัญหาแบบ N กลุ่ม ระบบจะประกอบด้วยตัวจำแนกแบบสองกลุ่มจำนวน N(N-1)/2 ตัว แต่ละตัวเป็นฟังก์ชันแบบทวิภาค โดยจะมีเพียง N-1 โนดที่เอดีเอจีเลือกมาใช้ โนดของเอดีเอจีถูกจัดเรียงเป็นรูปสามเหลี่ยมคว่ำ โดยมีโนดจำนวน N/2 โนด(ปัดขึ้น) ที่ระดับบนสุด และจำนวนโนดจะลดลงครึ่งหนึ่งสำหรับระดับถัดมา จนเหลือโนดเดียวในระดับล่างสุด ดังแสดงในรูปที่ 8



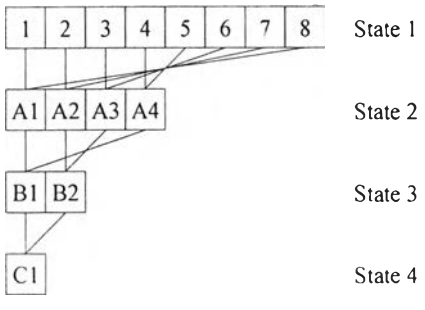
รูปที่ 8 โครงสร้างของเอดีเอจี ซึ่งประกอบด้วยโนด 7 โนดสำหรับปัญหาแบบ 8 กลุ่ม

ในการจำแนกโดยใช้เอดีเอจี ตัวจำแนกแบบสองกลุ่มในแต่ละโนดจะกำจัดกลุ่มหนึ่งออกไปโดยเริ่มจากระดับบนสุด และส่งข้อความที่บ่งถึงกลุ่มที่เลือกมายังโนดระดับล่าง ทำให้ในแต่ละรอบ จำนวนกลุ่มที่มีโอกาสถูกเลือกว่าเป็นกลุ่มที่ถูกต้องจะลดลงทีละครึ่ง การเลือกตัวจำแนกแบบสองกลุ่มในโนดแต่ละโนดของระดับถัดมา จะเลือกจากข้อความที่ส่งมาจากโนดระดับบนสองโนด (หนึ่งโนดในกรณีที่มาจากโนดกลางของระดับบน โดยโนดระดับบนมีจำนวนตัวจำแนกเป็นเลขคี่) กระบวนการนี้จะดำเนินไปอย่างต่อเนื่องจนเหลือโนดเดียวที่ระดับล่างสุด

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เมื่อสิ้นสุดกระบวนการ ผลการจำแนกจะมาจากข้อความที่ส่งมาจากโนดล่างสุด เอดีเอจีใช้โนดทั้งหมด N-1 โนดในการจำแนกเช่นเดียวกับดีดีเอจี สังเกตว่ากลุ่มที่ถูกต้องจะถูกจำแนกทั้งหมดอย่างมาก Log_2N ครั้งซึ่งต่ำกว่าจำนวนที่ต้องการในโครงสร้างแบบดีดีเอจี ซึ่งจำนวนครั้งแปรตามค่า N

ในทางปฏิบัติ การจำแนกแบบหลายกลุ่มโดยใช้เอดีเอจีสามารถทำได้โดยใช้ลิสต์เช่นเดียวกับดีดีเอจี รูปที่ 9 แสดงลิสต์สำหรับกรณี 8 กลุ่ม เริ่มแรกลิสต์จะประกอบด้วยสมาชิกจากกลุ่มทุกกลุ่มคือ 1 ถึง 8 ข้อมูลที่ต้องการจำแนกจะถูกจำแนกด้วยตัวจำแนกแบบสองกลุ่ม ที่เป็นคู่ของสมาชิกตัวแรกและตัวสุดท้ายในลิสต์ ซึ่งคือ 1vs8 และเมื่อตัวจำแนกเลือกกลุ่มใดแล้ว กลุ่มนั้นจะถูกส่งลงไปบรรจุในอีกลิสต์สำหรับระดับล่าง โดยใส่จากซ้ายไปขวา แล้วสมาชิกทั้งสองจะถูกลบออกจากลิสต์ระดับบน ต่อจากนั้นทำซ้ำขั้นตอนเดิมในลิสต์ระดับบน โดยตัวจำแนกแบบสองกลุ่มตัวถัดมาจะเป็น 2vs7 3vs6 และ 4vs5 ตามลำดับ ซึ่งกลุ่มที่ไม่ถูกกำจัดออกจะถูกนำไปใส่ยังตำแหน่งที่ 2 3 และ 4 ของลิสต์ในระดับที่สองต่อไป เป็นอันสิ้นสุด 1 รอบ ในแต่ละรอบกระบวนการจำแนกแบบสองกลุ่มนี้จะดำเนินต่อไปจนไม่เหลือสมาชิกหรือเหลือสมาชิก 1 ตัวในลิสต์ (ซึ่งจะถูกส่งลงมาในลิสต์ระดับถัดไปทันที เปรียบได้กับการชนะผ่าน) ในแต่ละรอบ ลิสต์ระดับถัดมาจะเหลือสมาชิกจำนวนครึ่งหนึ่ง (ปัดขึ้น) กระบวนนี้จึงทำซ้ำทั้งหมด $Log_2 8$ หรือ 3 รอบก่อนจะเหลือสมาชิกตัวเดียวในลิสต์ ซึ่งสมาชิกตัวนั้นคือกลุ่มที่ได้รับการจำแนก



รูปที่ 9 วิธีการนำเอดีเอจีไปใช้ในทางปฏิบัติ

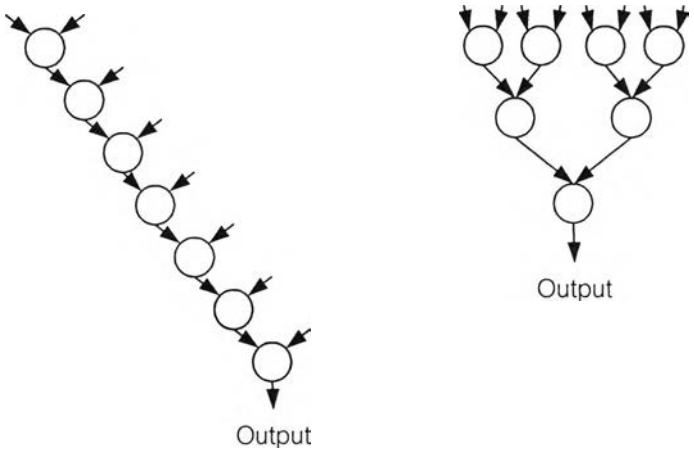
เอดีเอจีสามารถลดปัญหาที่เกิดกับดีดีเอจีในเรื่องการขึ้นกับลำดับของสมาชิกในลิสต์มาก โดยในกรณีปัญหาแบบ 4 กลุ่ม จะมีลิสต์ที่เป็นไปได้ 24 แบบ แต่มีเพียง 3 แบบเท่านั้นที่อาจให้ผลการจำแนกที่แตกต่างกันได้ (เทียบกับ 12 แบบในกรณีของดีดีเอจี) ลองตัวอย่างในตารางที่ 3 พบว่าเอดีเอจีให้ผลการจำแนกที่ไม่แปรเปลี่ยนไปมากเมื่อเปลี่ยนลำดับของสมาชิกในลิสต์ โดยจะยังคงให้คำตอบเป็นกลุ่ม 1 หรือ กลุ่ม 3 ทั้งหมดดังตารางที่ 4

ตารางที่ 4 โอกาสที่จะจำแนกจุด 'x' ในรูปที่ 7 เป็นกลุ่มต่างๆ โดยลิสต์ที่มีสมาชิก 4 ตัว

ผลการจำแนก	DDAG	ADAG
กลุ่ม 1	41.67%	66.67%
กลุ่ม 2	25.00%	-
กลุ่ม 3	25.00%	33.33%
กลุ่ม 4	8.33%	-

นอกจากนี้ความลึกของเส้นทางจำแนกของเอดีเอจียังลดลงต่ำกว่าของดีดีเอจี คืออยู่ในระดับ O(log N) เมื่อเทียบกับระดับ O(N) ของดีดีเอจี การลดลงของความลึกหมายถึงการลดลงของโอกาสที่จะจำแนกผิดพลาด ซึ่งจะทำให้ค่าคาดหวังของความถูกต้องสูงขึ้นในที่สุด

รูปที่ 10 แสดงให้เห็นเส้นทางของการจำแนกของดีดีเอจีและของเอดีเอจี สำหรับปัญหา 8 กลุ่ม แบบดีดีเอจีจะมีความลึกของเส้นทางตั้งแต่ 1 ถึง 7 ขึ้นกับว่ากลุ่มที่ถูกต้องเริ่มถูกเปรียบเทียบตั้งแต่ระดับใด ส่วนแบบเอดีเอจีจะมีความลึกของเส้นทางเป็น 3 ความแตกต่างของความลึกนี้จะเพิ่มขึ้นเมื่อจำนวนกลุ่มมากขึ้น



รูปที่ 10 รูปแสดงความลึกของโครงสร้างดีดีเอจีและเอดีเอจีตามลำดับ

ในส่วนถัดไป จะเสนอแบบจำลองที่ใช้ในการหาค่าคาดหวังของความถูกต้องของทั้งสองโครงสร้างเปรียบเทียบกัน พร้อมทั้งเสนอทฤษฎีค่าคาดหวังของความถูกต้องและบทพิสูจน์

3.3 การคำนวณหาค่าคาดหวังของความถูกต้องในแบบจำลองดีดีเอจี

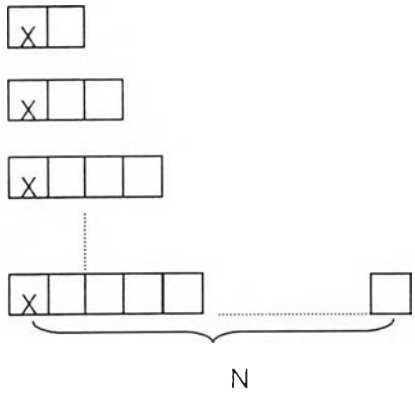
ระดับชั้นโดยเฉลี่ยที่ลึกกว่าของดีดีเอจีส่งผลกระทบต่อความถูกต้องของการจำแนก เพื่อวิเคราะห์ถึงผลกระทบดังกล่าว จึงจำเป็นที่จะต้องสร้างโมเดลขึ้นมา เพื่อคำนวณหาค่าความถูกต้องคาดหวังของวิธีดีดีเอจี

ในโมเดลดีดีเอจี สำหรับปัญหา N กลุ่ม จะใช้ลิสต์ขนาด N สมาชิก สมาชิกแต่ละตัวแทนกลุ่มแต่ละกลุ่ม ในแต่ละรอบของการเปรียบเทียบ สมาชิกตัวแรกสุดกับตัวหลังสุดตัวใดตัวหนึ่งจะถูกกำจัดออกจากลิสต์ ทำให้ขนาดของลิสต์ลดลงทีละหนึ่ง จนเหลือสมาชิก 1 ตัว ซึ่งเป็นกลุ่มที่ถูกจำแนก

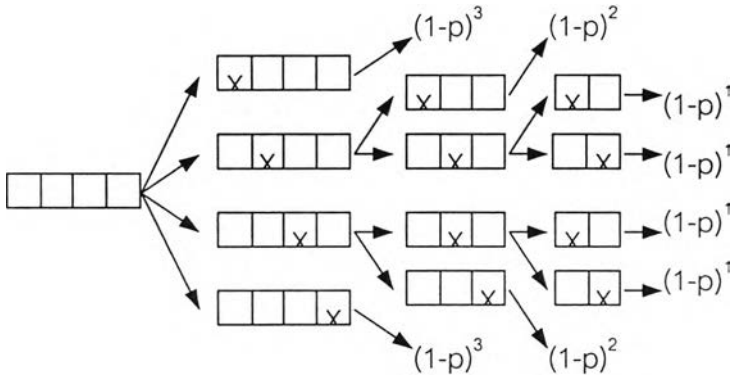
ข้อตกลงพื้นฐานของแบบจำลอง มีดังต่อไปนี้

1. ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะอยู่ในตำแหน่งต่างๆ ของลิสต์มีการแจกแจงแบบยูนิฟอร์ม
2. ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะถูกกำจัดออกอย่างผิดพลาดเป็น p เมื่อมีการเปรียบเทียบระหว่างกลุ่มที่ถูกต้องกับกลุ่มอื่น
3. ความน่าจะเป็นที่กลุ่มที่กลุ่มใดๆ ที่ไม่ใช่กลุ่มที่ถูกต้องจะถูกกำจัดออกเป็น 0.5 เมื่อมีการเปรียบเทียบระหว่างกลุ่ม 2 กลุ่มที่ไม่ใช่กลุ่มที่ถูกต้อง

แนวคิดหลักคือการจำแนกจะถูกต้องเมื่อกลุ่มที่ถูกต้องไม่ถูกกำจัดออกจากลิสต์ นั่นก็หมายความว่าเมื่อไรก็ตามที่กลุ่มที่ถูกต้องอยู่ตำแหน่งริม (ตัวแรกหรือตัวสุดท้ายของลิสต์) กลุ่มอื่นที่เหลือในลิสต์จะต้องเป็นฝ่ายถูกกำจัดออกจากลิสต์จนหมด และการแบ่งเหตุการณ์ที่จำแนกถูกเป็น N – 1 เหตุการณ์ ได้แก่เหตุการณ์ที่มีสมาชิกเหลือในลิสต์ 2,3,4,…,N ตัว ในขณะที่กลุ่มที่ถูกต้อง (แทนด้วย X) อยู่ที่ตำแหน่งริม*เป็นครั้งแรก* ดังรูปที่ 11



รูปที่ 11 เหตุการณ์ทั้งหมดที่เป็นไปได้ในการจำแนกที่ถูกต้อง



รูปที่ 12 ตัวอย่างปัญหาแบบ 4 กลุ่ม

วิธีคำนวณสามารถแสดงด้วยตัวอย่างการจำแนกปัญหาแบบ 4 กลุ่มดังรูปที่ 12 ภายใต้การแจกแจงแบบยูนิฟอร์ม ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะอยู่ในตำแหน่งต่างๆในลิสต์เป็น 1/4 ในกรณีที่กลุ่มที่ถูกต้องอยู่ที่ตำแหน่งริมของลิสต์ กลุ่มนั้นจะได้รับการจำแนกถูกต้องเมื่อกลุ่มอื่นทุกกลุ่มถูกกำจัดออกจากลิสต์ ซึ่งความน่าจะเป็นของเหตุการณ์นี้เป็น $(1-p)^{M-1}$ เมื่อ M เป็นจำนวนสมาชิกที่เหลือในลิสต์ขณะนั้น ในกรณีที่กลุ่มที่ถูกต้องไม่ได้อยู่ที่ตำแหน่งริม จะมีสองทางคือกลุ่มที่อยู่ตำแหน่งแรกจะถูกกำจัด หรือกลุ่มที่อยู่ตำแหน่งสุดท้ายจะถูกกำจัด ทำให้จำนวนของสมาชิกในลิสต์ลดลงทีละหนึ่ง จากตัวอย่างข้างต้น ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะได้รับการจำแนกอย่างถูกต้องเป็น

$$= \frac{1}{4}(1-p)^3 + \frac{1}{4}\left(\frac{1}{2}\right)^1(1-p)^2 + \frac{4}{4}\left(\frac{1}{2}\right)^2(1-p)^1 + \frac{1}{4}\left(\frac{1}{2}\right)^1(1-p)^2 + \frac{1}{4}(1-p)^3$$

$$= \frac{2}{4}(1-p)^3 + \frac{2}{4}\left(\frac{1}{2}\right)^1(1-p)^2 + \frac{4}{4}\left(\frac{1}{2}\right)^2(1-p)^1$$

$$= \frac{2}{4}(1-p)^3 + \frac{1}{4}(1-p)^2 + \frac{1}{4}(1-p)^1$$

$$= \frac{1}{4}\left[\frac{(1-p)(1-(1-p)^3)}{p} + (1-p)^3\right]$$

$$= \frac{1}{4}\left[\frac{(1-p)}{p} + (1-p)^3 - \frac{(1-p)^4}{p}\right]$$

ทฤษฎีบทที่ 1 ให้ p เป็นความน่าจะเป็นที่กลุ่มที่ถูกต้องจะถูกกำจัดออกจากลิสต์อย่างผิดพลาดเมื่อมีการเปรียบเทียบระหว่างกลุ่มที่ถูกต้องกับกลุ่มอื่น และให้ความน่าจะเป็นที่กลุ่มใดกลุ่มหนึ่งจากสองกลุ่มที่เปรียบเทียบกันจะถูกกำจัดออกเป็น 0.5 เมื่อทั้งสองกลุ่มไม่ใช่กลุ่มที่ถูกต้อง และ N เป็นจำนวนกลุ่ม จะได้ว่าในกรณีที่การแจกแจงความน่าจะเป็นของโอกาสที่กลุ่มที่ถูกต้องจะอยู่ในตำแหน่งใดๆของลิสต์เป็นแบบยูนิฟอร์ม ค่าคาดหวังของความถูกต้องของดีดีเอจีจะเป็น

$$\frac{1}{N}\left[\frac{1-p}{p} + (1-p)^{N-1} - \frac{(1-p)^N}{p}\right]$$

บทพิสูจน์ จากที่ปรากฏในตัวอย่างข้างต้น กลุ่มที่ถูกต้องจะถูกจำแนกอย่างถูกต้องถ้าเมื่อไรก็ตามที่กลุ่มที่ถูกต้องอยู่ที่ตำแหน่งริมของลิสต์ กลุ่มอื่นจะต้องถูกกำจัดออกจากลิสต์จนหมด

พิจารณากรณีของลิสต์ที่มีสมาชิก i ตัว เมื่อ $2 \le i \le N-1$ ในขณะที่กลุ่มที่ถูกต้องอยู่ตำแหน่งซ้ายสุด*เป็นครั้งแรก* ลิสต์ดังกล่าวสามารถเขียนในรูป XO...O เมื่อ X แทนกลุ่มที่ถูกต้องและ O แทนกลุ่มอื่นๆ ที่ไม่ถูกต้อง

ลิสต์นี้จะมาจากลิสต์ที่มีสมาชิกจำนวน i+1 ตัว โดยมีกลุ่มที่ไม่ใช่กลุ่มที่ถูกต้อง 1 ตัวอยู่ทางซ้ายของกลุ่มที่ถูกต้อง ซึ่งแสดงได้เป็น OXO...O ก่อนที่จะได้ลิสต์ OXO...O ได้มีการกำจัดกลุ่มที่ไม่ถูกต้องจำนวน N - i - 1 ตัวออกจากลิสต์ ดังนั้นถ้าเริ่มจากลิสต์เริ่มต้นทุกแบบจำนวนเส้นทางที่เป็นไปได้ทั้งหมดที่จะมาถึงจุดที่ลิสต์มีสมาชิก i ตัว โดยที่กลุ่มที่ถูกต้องอยู่ตำแหน่งซ้ายสุดจะเป็น 2^{N-i-1} (เนื่องจากมีสองทางเลือกที่เป็นไปได้คือจะกำจัดกลุ่มซ้ายสุดหรือกำจัดกลุ่มขวาสุดออกจากลิสต์) ดังนั้นจำนวนเส้นทางที่เป็นไปได้ทั้งหมดที่มาถึงลิสต์ที่มีสมาชิก i ตัวโดยที่กลุ่มที่ถูกต้องอยู่ที่ตำแหน่งริมจึงเป็น 2^{N-i} ความน่าจะเป็นที่จะได้ลิสต์ที่มีสมาชิก i ตัวโดยที่กลุ่มที่ถูกต้องอยู่ที่ตำแหน่งริมจึงเป็น

$$\left(\frac{1}{N}\right)\left(\frac{1}{2}\right)^{N-i}(1-p)^{i-1}$$

เนื่องจากกลุ่มที่ไม่ถูกต้อง i-1 กลุ่มจะต้องถูกกำจัดออกเมื่อเปรียบเทียบกับกลุ่มที่ถูกต้อง ความน่าจะเป็นรวมที่กลุ่มที่ถูกต้องในรูปแบบนี้จะถูกจำแนกอย่างถูกต้องจึงเป็น

$$2^{N-i}\left(\frac{1}{N}\right)\left(\frac{1}{2}\right)^{N-i}(1-p)^{i-1} = \left(\frac{1}{N}\right)(1-p)^{i-1}$$

จากนี้จึงพิจารณากรณีที่กลุ่มที่ถูกต้องอยู่ที่ตำแหน่งริมเป็นครั้งแรกเมื่อลิสต์มีสมาชิก N ตัว สำหรับกรณีนี้จะเห็นได้ชัดว่าความน่าจะเป็นที่กลุ่มที่ถูกต้องจะถูกจำแนกอย่างถูกต้องจะเป็น

$$2\left(\frac{1}{N}\right)(1-p)^{N-1}$$

สุดท้ายจึงรวมความน่าจะเป็นทั้งหมดเข้าด้วยกัน แล้วจะได้ค่าคาดหวังของความถูกต้อง ดังนี้

$$\begin{aligned}
&= \sum_{i=2}^{N-1}\left(\frac{1}{N}(1-p)^{i-1}\right)+\frac{2}{N}(1-p)^{N-1} \\
&= \sum_{j=1}^{N-1}\left(\frac{1}{N}(1-p)^{j}\right)+\frac{1}{N}(1-p)^{N-1} \\
&= \frac{1}{N}\left[\frac{(1-p)\left(1-(1-p)^{N-1}\right)}{p}+(1-p)^{N-1}\right] \\
&= \frac{1}{N}\left[\frac{1-p}{p}+(1-p)^{N-1}-\frac{(1-p)^{N}}{p}\right]
\end{aligned}$$

QED

3.4 การคำนวณหาค่าคาดหวังของความถูกต้องในแบบจำลองเอดีเอจี

ระดับชั้นโดยเฉลี่ยที่ลึกกว่าของดีดีเอจีส่งผลกระทบต่อความถูกต้องของการจำแนก เพื่อวิเคราะห์ถึงผลกระทบดังกล่าว จึงจำเป็นที่จะต้องสร้างโมเดลขึ้นมา เพื่อคำนวณหาค่าความถูกต้องคาดหวังของวิธีเอดีเอจี

ในโมเดลดีดีเอจี สำหรับปัญหา N กลุ่ม จะใช้ลิสต์ขนาด N สมาชิก สมาชิกแต่ละตัวแทนแต่ละกลุ่ม ในแต่ละรอบของการเปรียบเทียบ จะทำการเปรียบเทียบเป็นคู่ สมาชิกตัวริม 2 ตัวจะถูกดึงออกจากลิสต์เพื่อเปรียบเทียบกัน และใส่ตัวที่ถูกเลือกในลิสต์ใหม่ที่มีขนาดเป็นครึ่งหนึ่ง ทำจนเหลือสมาชิก 1 ตัว ซึ่งจะเป็นกลุ่มที่ถูกจำแนก

ข้อตกลงพื้นฐานของแบบจำลอง มีดังต่อไปนี้

1. ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะอยู่ในตำแหน่งต่างๆ ของลิสต์มีการแจกแจงแบบยูนิฟอร์ม
2. ความน่าจะเป็นที่กลุ่มที่ถูกต้องจะถูกกำจัดออกอย่างผิดพลาดเป็น p เมื่อมีการเปรียบเทียบระหว่างกลุ่มที่ถูกต้องกับกลุ่มอื่น
3. ความน่าจะเป็นที่กลุ่มที่กลุ่มใดๆ ที่ไม่ใช่กลุ่มที่ถูกต้องจะถูกกำจัดออกเป็น 0.5 เมื่อมีการเปรียบเทียบระหว่างกลุ่ม 2 กลุ่มที่ไม่ใช่กลุ่มที่ถูกต้อง

ทฤษฎีบทที่ 2 ให้ p เป็นความน่าจะเป็นที่กลุ่มที่ถูกต้องจะถูกกำจัดออกจากลิสต์อย่างผิดพลาดเมื่อมีการเปรียบเทียบระหว่างกลุ่มที่ถูกต้องกับกลุ่มอื่น และให้ความน่าจะเป็นที่กลุ่มใดกลุ่มหนึ่งจากสองกลุ่มที่เปรียบเทียบกันจะถูกกำจัดออกจากลิสต์เป็น 0.5 เมื่อทั้งสองกลุ่มนั้นไม่ใช่กลุ่มที่ถูกต้อง จะได้ว่าในกรณีที่การแจกแจงความน่าจะเป็นของโอกาสที่กลุ่มที่ถูกต้องจะอยู่ในตำแหน่งใดๆของลิสต์เป็นแบบยูนิฟอร์ม ค่าคาดหวังของความถูกต้องของเอดีเอจีจะเป็น

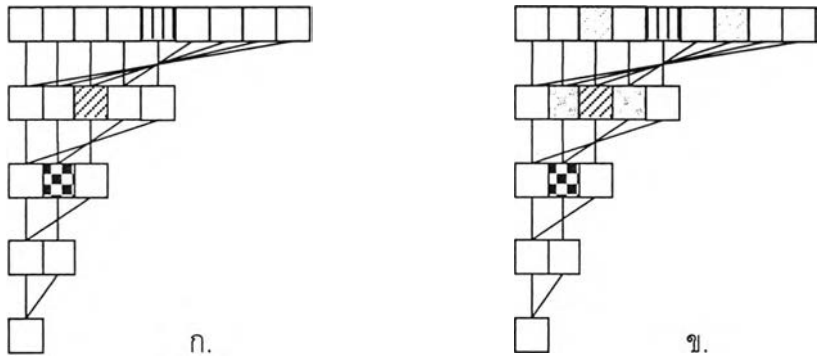
$$\left(\frac{2N - 2^{\lceil \log_2 N \rceil}}{N}\right)(1-p)^{\lceil \log_2 N \rceil} + \left(\frac{2^{\lceil \log_2 N \rceil} - N}{N}\right)(1-p)^{\lceil \log_2 N \rceil - 1}$$

เมื่อ N เป็นจำนวนกลุ่ม และ $\lceil x \rceil$ เป็นเลขจำนวนเต็มที่น้อยที่สุดที่มากกว่าหรือเท่ากับ x

บทพิสูจน์ ในกรณีปัญหาแบบ N กลุ่ม ความสูงหรือจำนวนของระดับทั้งหมดของเอดีเอจีเป็น $\lceil \mathrm{Log}_2 N \rceil$ และในการจำแนกที่ถูกต้องจำนวนครั้งของการเปรียบเทียบขึ้นกับตำแหน่งเริ่มต้นของกลุ่มที่ถูกต้อง ในบางตำแหน่งกลุ่มที่ถูกต้องจะถูกเปรียบเทียบกับกลุ่มอื่นจำนวน $\lceil \mathrm{Log}_2 N \rceil$ ครั้ง ในขณะที่ในบางตำแหน่งกลุ่มที่ถูกต้องจะถูกเปรียบเทียบน้อยกว่า $\lceil \mathrm{Log}_2 N \rceil$ ครั้ง เรียกกลุ่มหลังนี้ว่า กลุ่มชนะผ่าน

เนื่องจากตามสถาปัตยกรรมของเอดีเอจี กลุ่มชนะผ่านเป็นกลุ่มที่อยู่ตรงตำแหน่งกลางของลิสต์ที่มีสมาชิกเป็นเลขคี่ และหลังจากผ่านลงสู่ระดับถัดมากลุ่มชนะผ่านเหล่านี้จะไปอยู่

ที่ตำแหน่งริม ดังนั้นกลุ่มชนะผ่านจึงชนะผ่านได้เพียงครั้งเดียว นั่นหมายความว่ากลุ่มชนะผ่านจะถูกเปรียบเทียบกับกลุ่มอื่น $\lceil Log_2 N \rceil - 1$ ครั้ง



รูปที่ 13 แสดงตำแหน่งของกลุ่มที่จะถูกจำแนกน้อยกว่ากลุ่มที่ตำแหน่งอื่น 1 ครั้ง

จำนวนตำแหน่งของกลุ่มชนะผ่านจะขึ้นอยู่กับจำนวนของกลุ่ม ถ้าจำนวนกลุ่มสามารถเขียนให้อยู่ในรูป 2^X เมื่อ X เป็นเลขจำนวนเต็มบวก จะไม่มีตำแหน่งที่สามารถชนะผ่าน แต่ถ้าจำนวนกลุ่มไม่สามารถเขียนให้อยู่ในรูป 2^X ได้ จะมีตำแหน่งที่สามารถชนะผ่าน

เราสามารถคำนวณหาจำนวนกลุ่มชนะผ่านนี้ได้ จากความรู้ที่ว่ากลุ่มชนะผ่านจะอยู่ในตำแหน่งกลางของลิสต์ในแต่ละระดับที่จำนวนสมาชิกเป็นเลขคี่ เช่นตำแหน่งที่ 5 ของลิสต์เริ่มต้น ตำแหน่งที่ 3 ของลิสต์ในระดับที่ 2 และตำแหน่งที่ 2 ของลิสต์ในระดับที่ 3 ดังรูปที่ 13 ก. กลุ่มชนะผ่านในระดับที่ i สามารถมาจากตำแหน่งใดตำแหน่งหนึ่งในสองตำแหน่งที่ระดับ i - 1 ดังในรูปที่ 13 ข. ซึ่งมีตำแหน่งในลิสต์เริ่มต้นของกลุ่มชนะผ่านทั้งหมด 7 ตำแหน่ง ให้ F(N) เป็นจำนวนของตำแหน่งทั้งหมดของกลุ่มชนะผ่านในลิสต์เริ่มต้นที่มีสมาชิก N ตัว เห็นได้ชัดว่า F(2) = 0 และ

$$F(N) = (N \bmod 2) + 2 \cdot F(\lceil N/2 \rceil). \quad (1)$$

จากนี้ จะพิสูจน์แบบอุปนัยบน X ว่า $F(N) = 2^{\lceil Log_2 N \rceil} - N$ โดยที่ $N \in I^+$, $2^X + 1 \leq N \leq 2^{X+1}$ และ $X \in I^+$, $X \geq 1$

อันดับแรก พิจารณากรณีฐานที่ X = 1 ในกรณีนี้ N เท่ากับ 3 ถึง 4 เห็นได้ชัดว่า $F(3)=1$ ซึ่งสอดคล้องกับ $F(3) = 2^{\lceil Log_2 3 \rceil} - 3$ และ $F(4)=0$ ซึ่งสอดคล้องกับ $F(4) = 2^{\lceil Log_2 4 \rceil} - 4$

จากนั้นจะพิสูจน์กรณีทั่วไปที่ $X \geq 2$ โดยวิธีอุปนัย ให้ $F(N) = 2^{\lceil Log_2 N \rceil} - N$ เมื่อ N เป็นเลขจำนวนเต็มระหว่าง $2^{X-1} + 1$ และ 2^X แล้วจะพิสูจน์ว่า F(M) ก็เท่ากับ $2^{\lceil Log_2 M \rceil} - M$ เช่นกัน เมื่อ M เป็นจำนวนเต็มระหว่าง 2^X+1 และ 2^{X+1}

ในกรณีที่ M = 2N (เลขคู่) และ $2^X+2 \le M \le 2^{X+1}$ เห็นได้ชัดว่า F(M) = 2F(N) ตามสมการที่ 1 (ในกรณีที่ M เป็นเลขคู่ กลุ่มชนะผ่านของ F(N) จะมาจาก กลุ่มชนะผ่านของ F(M) 2 ตำแหน่ง) ดังนั้นจะได้

$$\begin{aligned} F(M) &= 2F(N) \\ &= 2(2^{\lceil \text{Log}_2 N \rceil} - N) \\ &= 2^{\lceil \text{Log}_2 2N \rceil} - 2N \\ &= 2^{\lceil \text{Log}_2 M \rceil} - M \end{aligned}$$

ซึ่งพิสูจน์กรณีแรกที่ M เป็นเลขคู่

จากนั้นในกรณีที่ M = 2N – 1(เลขคี่) และ $2^X+1 \le M \le 2^{X+1}- 1$ เช่นเดิมจะพบว่า F(M) = 2F(N) + 1 จากสมการที่ 1 (ในกรณีที่ M เป็นเลขคี่ จำนวนตำแหน่งชนะผ่านของ F(N) จะมาจาก ตำแหน่งชนะผ่านสองตำแหน่งของ F(M) และยังสมาชิกมีอีก 1 ตัวของ F(M) ที่ชนะผ่านในรอบนี้) ดังนั้นจะได้

$$\begin{aligned} F(M) &= 2F(N) + 1 \\ &= 2(2^{\lceil \text{Log}_2 N \rceil} - N) + 1 \\ &= 2^{\lceil \text{Log}_2 N \rceil + 1} - 2N + 1 \\ &= 2^{\lceil \text{Log}_2 2N \rceil} - (2N - 1) \end{aligned}$$

เนื่องจาก $2^X+1 \le 2N - 1 < 2N \le 2^{X+1}$ จะได้ว่า $\log_2(2^X+1) \le \log_2(2N – 1) < \log_2(2N) \le \log_2(2^{X+1})$ ซึ่งหมายความว่า $\lceil \log_2(2N – 1) \rceil = \lceil \log_2(2N) \rceil = X+1$ สูตรข้างต้นจึงเป็น

$$\begin{aligned} F(M) &= 2^{\lceil \text{Log}_2 (2N-1) \rceil} - (2N - 1) \\ &= 2^{\lceil \text{Log}_2 M \rceil} - M \end{aligned}$$

ซึ่งพิสูจน์กรณี M เป็นเลขคี่

ทั้งหมดนี้จึงพิสูจน์ว่า F(N) เท่ากับ $2^{\lceil \text{Log}_2 N \rceil} - N$ เมื่อ N เป็นจำนวนเต็มที่อยู่ระหว่าง 2^X+1 กับ $2^{X+1}- 1$ เมื่อ X มากกว่าหรือเท่ากับ 1

เมื่อได้สูตรของค่า F(N) แล้ว จะสามารถหาค่าคาดหวังของความถูกต้องของเอดีเอจีได้ เนื่องจากความน่าจะเป็นที่กลุ่มที่ถูกต้องจะอยู่ที่ตำแหน่งต่างๆเป็น 1/N ภายใต้การแจกแจงแบบยูนิฟอร์ม จะได้ว่าค่าคาดหวังของความถูกต้องสามารถหาได้จากการถ่วงน้ำหนักของความถูกต้องของกลุ่มชนะผ่านด้วย F(N)/N และถ่วงน้ำหนักของกลุ่มที่ไม่ชนะบายด้วย (N-F(N))/N สุดท้ายจะได้ค่าคาดหวังของความถูกต้องเป็น

$$\begin{aligned} &= \frac{N - F(N)}{N}(1-p)^{\lceil \log_2 N \rceil} + \frac{F(N)}{N}(1-p)^{\lceil \log_2 N \rceil - 1} \\ &= \left(\frac{2N - 2^{\lceil \log_2 N \rceil}}{N}\right)(1-p)^{\lceil \log_2 N \rceil} + \left(\frac{2^{\lceil \log_2 N \rceil} - N}{N}\right)(1-p)^{\lceil \log_2 N \rceil - 1} \end{aligned}$$

QED

ตารางที่ 5 ตัวอย่างของความลึกของเส้นทางในตำแหน่งเริ่มต้นแต่ละตำแหน่งของลิสต์

จำนวนชั้น	N	F(N) $= 2^{\lceil \log_2 N \rceil} - N$		จำนวนครั้งของการเปรียบเทียบที่ตำแหน่งต่างๆ
1	2	$2^1 - 2$	0	11
2	3	$2^2 - 3$	1	212
	4	$2^2 - 4$	0	2222
3	5	$2^3 - 5$	3	32223
	6	$2^3 - 6$	2	323323
	7	$2^3 - 7$	1	3332333
	8	$2^3 - 8$	0	33333333
4	9	$2^4 - 9$	7	433333334
	10	$2^4 - 10$	6	4333443334
	11	$2^4 - 11$	5	43443334434
	12	$2^4 - 12$	4	434434434434
	13	$2^4 - 13$	3	4443443443444
	14	$2^4 - 14$	2	44434444443444
	15	$2^4 - 15$	1	444444434444444
	16	$2^4 - 16$	0	4444444444444444

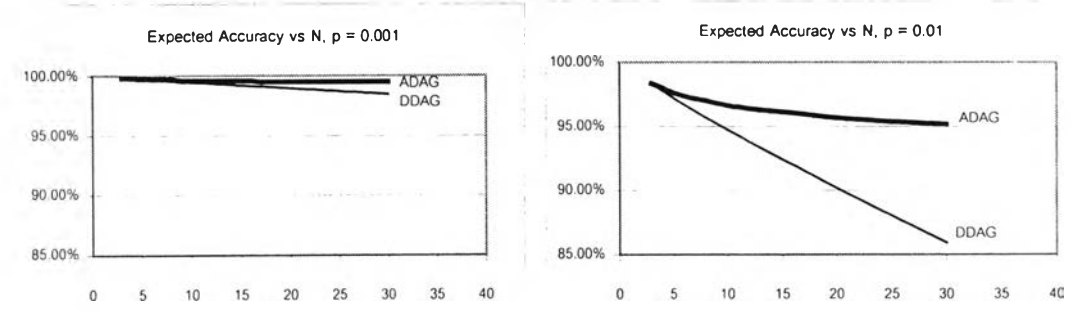
3.5 การวิเคราะห์โครงสร้างของดีดีเอจีเทียบกับเอดีเอจี

จากทฤษฎีพบว่าความถูกต้องของวิธีทั้งสองวิธีตามแบบจำลองเป็นดังตารางที่ 6

ตารางที่ 6 เปรียบเทียบค่าความถูกต้องสำหรับกรณีดีสุด กรณีเฉลี่ย และกรณีเลวสุด

กรณีดีสุด	ADAG	$(1-p)^{\lceil \log_2 N \rceil - 1}$
	DDAG	$(1-p)$
กรณีเฉลี่ย	ADAG	$\left(\frac{2N - 2^{\lceil \log_2 N \rceil}}{N}\right)(1-p)^{\lceil \log_2 N \rceil} + \left(\frac{2^{\lceil \log_2 N \rceil} - N}{N}\right)(1-p)^{\lceil \log_2 N \rceil - 1}$
	DDAG	$\frac{1}{N}\left[\frac{1-p}{p} + (1-p)^{N-1} - \frac{(1-p)^N}{p}\right]$
กรณีเลวสุด	ADAG	$(1-p)^{\lceil \log_2 N \rceil}$
	DDAG	$(1-p)^N$

พบว่าค่าความถูกต้องโดยเฉลี่ยจะสัมพันธ์เชิงบวกกับจำนวนกลุ่ม (N) และอัตราการผิดพลาด (p) ที่ตัวจำแนกแบบสองกลุ่ม กล่าวคือเมื่อจำนวนกลุ่มเพิ่มขึ้นจะพบว่าค่าความถูกต้องลดลงในทั้งสองขั้นตอนวิธี แต่ในแบบ เอดีเอจีจะลดลงในอัตราที่ช้ากว่า ดังรูปที่ 14



รูปที่ 14 กราฟแสดงความสัมพันธ์ระหว่างจำนวนกลุ่มกับค่าคาดหวังของความถูกต้อง ในกรณีที่อัตราการผิดพลาดเป็น 0.001 และ 0.01 ตามลำดับ