บทที่ 3

ฟังก์ชันถ่วงน้ำหนักสำหรับอัลกอริทึมเนียเรสท์เนเบอร์

ในบทนี้ เราจะเริ่มต้นกระบวนการของวิทยานิพนธ์ฉบับนี้ในส่วนทฤษฎีโดยการนิยามคุณสมบัติของฟังก์ชันถ่วงน้ำหนักที่ต้องการเพื่อหารูปแบบของฟังก์ชันถ่วงน้ำหนักที่เหมาะสม

3.1 **คุณสมบัติเบื้องต้นของฟังก์ชันถ่วงน้ำหนักที่ต้องการ**

หลังจากที่ได้ทำการพิจารณาเกี่ยวกับฟังก์ชันระยะทางรูปแบบต่างๆ เราได้เลือกที่จะใช้ระยะทางยูคลิดระหว่างจุดพิกัดทั้งสองที่สนใจมาใช้เป็นอินพุทของฟังก์ชันนี้โดยไม่สนใจค่าของจุดพิกัดทั้งสอง โดยสาเหตุที่เลือกใช้ระยะทางยูคลิดนั้นก็เนื่องมาจากระยะทางชนิดนี้น่าจะสะท้อนถึงความสัมพันธ์กันของข้อมูลต่างๆ ได้ดีในทางคณิตศาสตร์

ด้วยเหตุนี้จากที่ได้เคยนิยามฟังก์ชันถ่วงน้ำหนักในรูปแบบทั่วไปที่สุดโดยสัญลักษณ์ $\mathrm{f}(p,d,D)$ ในการอธิบายอัลกอริทึมจำพวกเนียเรสท์เนเบอร์ โดยที่ p เป็นข้อมูลที่ต้องการจำแนกประเภท d เป็นข้อมูลสอนที่กำลังพิจารณา และ D เป็นเซตของข้อมูลสอนทั้งหมด ด้วยข้อกำหนดเบื้องต้นในการเลือกที่จะใช้ระยะทางยูคลิดเพียงอย่างเดียวและทำการพิจารณาเฉพาะกรณีที่ผลลัพธ์ของฟังก์ชันนี้จะไม่เกี่ยวข้องกับข้อมูลตัวอื่นนอกจากข้อมูลทั้งสองที่กำลังสนใจอยู่ เราจึงสามารถเปลี่ยนรูปของฟังก์ชัน $\mathrm{f}(p,d,D)$ ให้ง่ายขึ้นได้เป็น $f(x)$ โดยฟังก์ชันนี้จะรับอินพุทค่าเดียวคือ x ซึ่งเป็นระยะทางยูคลิดระหว่างจุด p และ d นั่นก็คือเป็นระยะทางระหว่างข้อมูลทั้งสองจุดที่กำลังสนใจ และถือว่า D ซึ่งเป็นข้อมูลสอนอื่นๆ ไม่มีผลต่อการพิจารณาค่าของฟังก์ชันในคู่นั้นๆ

เมื่อทำการพิจารณาฟังก์ชันต่างๆ ที่เป็นไปตามข้อกำหนดกล่าวถึง เราจะสามารถนิยามคุณสมบัติทั่วไปของฟังก์ชันถ่วงน้ำหนักได้ดังนี้

1. ฟังก์ชันนี้จะรับอินพุทเพียงค่าเดียวซึ่งก็คือระยะทางยูคลิดระหว่างจุด 2 จุด เราจะเรียกฟังก์ชันนี้ว่า $f(x)$

2. ฟังก์ชันนี้มีโดเมนคือ $[0,\infty)$ และเรนจ์ $[0,\infty)$ เนื่องจากระยะทางและคะแนนของผลลัพธ์จะต้องไม่ติดลบ

3. $f(x)$ เป็นฟังก์ชันต่อเนื่องชนิดไม่เพิ่ม เนื่องมาจากข้อมูลที่มีความแตกต่างจากข้อมูลที่ต้องการจำแนกประเภทมากๆ ควรจะมีผลต่อผลลัพธ์ของการจำแนกประเภทไม่มากไปกว่าข้อมูลที่มีคล้ายคลึงกับข้อมูลที่ต้องการจำแนกประเภท และคะแนนที่ได้ควรจะเปลี่ยนแปลงอย่างสม่ำเสมอเมื่อเทียบกับอินพุทของฟังก์ชัน

4. $f(0)=\infty$ เนื่องจากข้อมูลที่มีคุณสมบัติตรงกันกับข้อมูลที่ต้องการจำแนกประเภททุกอย่างจะมีผลเพียงพอต่อการจำแนกประเภทแล้ว และ $\lim_{x\to\infty} f(x)=0$ เนื่องจากข้อมูลที่แตกต่างกันมากจนถึงขั้นที่ไม่มีความเกี่ยวข้องการเลยไม่ควรจะมีผลต่อการจำแนกประเภท ไม่ว่าจะเป็นในด้านการสนับสนุนหรือการคัดค้านผลลัพธ์เหล่านั้นก็ตาม

3.2 การทนทานต่อการสเกล

คุณสมบัติที่ดีอย่างหนึ่งที่ไม่ใช่คุณสมบัติทั่วไปก็คือฟังก์ชันนี้ควรจะมีค่าเท่าเดิมแม้จะมีการปรับสเกลของคุณสมบัติ นั่นก็คือเมื่อคุณสมบัติทั้งหมดถูกปรับสเกลในเชิงเส้นแบบเท่าเทียมกัน อัลกอริทึมเนียเรสท์เนเบอร์ที่ใช้ฟังก์ชันนี้ควรจะยังให้ผลลัพธ์ของการจำแนกประเภทเช่นเดิม

ทฤษฎีบทที่ 1: ฟังก์ชันที่ต้องการจะอยู่ในรูปแบบ $f(x)=\frac{1}{x^k}$ เมื่อ k เป็นจำนวนจริงบวก

พิสูจน์: จากคุณสมบัติในการทนทานต่อการสเกลของฟังก์ชัน เราสามารถสรุปได้ว่าสำหรับระยะทาง r_1 และ r_2 ใดๆ และจำนวนเต็มบวก a ใดๆ ซึ่งเป็นอัตราส่วนของการสเกล

$$\frac{f(ar_1)}{f(ar_2)}=\frac{f(r_1)}{f(r_2)}$$

จากสมการนี้เราสามารถที่จะสรุปได้ว่า สำหรับระยะทาง r ใดๆ และแต่ละอัตราส่วนการสเกล a เราจะสามารถหาค่าคงที่ α ได้ค่าหนึ่งซึ่ง $\alpha=\frac{f(ar)}{f(r)}$

คราวนี้ลองพิจารณาจำนวนจริงที่ไม่เป็นลบ a หรือ b ใดๆ เราจะได้ว่า

$$\begin{aligned} f(ab) &= \alpha f(b) \\ &= \frac{f(a)f(b)}{f(1)} \end{aligned}$$

ให้ $h(\ln(x))=\ln\left(\frac{f(x)}{f(1)}\right)$

$$f(ab)=\frac{f(a)f(b)}{f(1)}$$

$$\frac{f(ab)}{f(1)}=\left(\frac{f(a)}{f(1)}\right)\left(\frac{f(b)}{f(1)}\right)$$

$$\ln\left(\frac{f(ab)}{f(1)}\right)=\ln\left(\frac{f(a)}{f(1)}\right)+\ln\left(\frac{f(b)}{f(1)}\right)$$

$$h(\ln(ab))=h(\ln(a))+h(\ln(b))$$

$$h(\ln(a)+\ln(b)) = h(\ln(a)) + h(\ln(b))$$

หลังจากแทนที่ $\ln(a)$ และ $\ln(b)$ ด้วย c และ d เราจะได้ว่า

$$h(c+d) = h(c) + h(d)$$

จากสมการสุดท้ายนี้จะเห็นได้ว่า $h(x)$ สามารถที่จะถูกแยกออกมาได้ด้วยการบวกดังนั้น จึงเป็นการชัดเจนแล้วว่ารูปแบบทั่วไปที่สุดของ $h(x)$ จะเป็น $h(x) = tx$ เมื่อ t เป็นจำนวนจริงใดๆ แล้วเราก็จะสามารถแก้สมการเพื่อหารูปแบบของฟังก์ชัน $f(x)$ ได้ดังต่อไปนี้

$$h(\ln(x)) = \ln\left(\frac{f(x)}{f(1)}\right)$$

$$t\ln(x) = \ln\left(\frac{f(x)}{f(1)}\right)$$

$$f(x) = f(1)e^{t\ln(x)}$$

$$f(x) = f(1)x^t$$

คราวนี้ให้ $m = f(1)$ แล้วเราก็จะได้สมการ

$$f(x) = mx^t$$

จากการที่ทั้งโดเมนและเรนจ์ของ $f(x)$ เป็นจำนวนจริงที่ไม่เป็นลบ เราสามารถสรุปได้ว่า m จะต้องเป็นจำนวนจริงบวก และจากคุณสมบัติที่ 3 และ 4 ของฟังก์ชัน เราจะได้ว่า $\lim_{x\to 0} f(x) = \infty$ จากสมการนี้และคุณสมบัติข้อที่ 4 ซึ่ง $\lim_{x\to\infty} f(x) = 0$ จะเป็นผลให้ $t < 0$ โดยกรณีที่ $x = 0$ ซึ่งจะให้ผลลัพธ์ที่ไม่นิยามจากการหารด้วย 0 จะต้องถูกจัดการเป็นกรณีพิเศษ

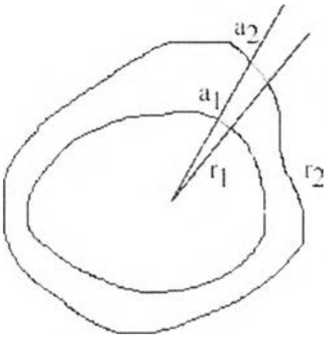
ในเมื่อการใช้ฟังก์ชันถ่วงน้ำหนักนั้นเป็นไปโดยการรวมผลที่ได้จากฟังก์ชันนั้นระหว่างข้อมูลที่ให้ผลลัพธ์เดียวกันด้วยการบวกแล้วเปรียบเทียบกับค่าจากข้อมูลที่ให้ผลลัพธ์อื่น ดังนั้นค่าของ m จึงจะไม่ส่งผลต่อผลลัพธ์ของการจำแนกประเภทตราบใดก็ตามที่ m เป็นจำนวนจริงบวก ดังนั้นโดยไม่สูญเสียลักษณะทั่วไป เราจะให้ $m = 1$ ซึ่งจะส่งผลให้สมการอยู่ในรูปแบบที่เรียบง่ายที่สุด จากนั้นจึงปรับรูปแบบของสมการใหม่ซึ่งจะทำให้เราได้ว่า

$$f(x) = \frac{1}{x^t}; t > 0$$

3.3 **ความมีประโยชน์ของข้อมูล**

ลองพิจารณาสถานการณ์ต่อไปนี้ สมมุติให้เรารู้ฟังก์ชันของผลลัพธ์ที่แท้จริงของข้อมูลใดๆ บนพื้นผิวของรูปทรง n มิติ ซึ่งมีข้อมูลที่เราต้องการจำแนกประเภทอยู่ข้างใน เมื่อ n เป็นจำนวน

ของคุณสมบัติของข้อมูล จากการนี้ การรู้ข้อมูลที่อยู่ภายนอกของรูปทรงนั้นจะไม่ช่วยเพิ่มข้อมูลที่เป็นประโยชน์ในการจำแนกประเภทของข้อมูลที่เราต้องการ นั่นก็คือสมมุติว่าเรารู้ฟังก์ชันของผลลัพธ์บนพื้นผิวของรูปทรงอีกอันที่มีรูปทรงอันแรกอยู่ข้างใน ฟังก์ชันของผลลัพธ์อันใหม่นี้ก็จะไม่เป็นประโยชน์ต่อการหาผลลัพธ์ของการจำแนกประเภทเช่นกัน อย่างไรก็ตาม ข้อมูลที่เป็นอินพุทที่แท้จริงในงานของเราจะไม่อยู่ในรูปแบบของฟังก์ชันเช่นนี้แต่จะอยู่ในรูปแบบของจุดหลายๆจุด ซึ่งจะส่งผลให้ข้อมูลทั้งหมดล้วนมีความสำคัญต่อผลลัพธ์ ดังนั้นถึงแม้ข้อมูลที่อยู่บนพื้นผิวของข้อมูลที่อยู่บนพื้นผิวของรูปทรงที่อยู่ภายนอกจะไม่ได้ไร้ประโยชน์โดยสิ้นเชิงแต่เราก็ยังสามารถบอกได้ว่าข้อมูลตัวนี้จะมีประโยชน์น้อยกว่าหรือมีประโยชน์ไม่มากกว่าข้อมูลที่อยู่ภายในรูปทรงนั้น นั่นคือ ข้อมูลตัวที่อยู่ด้านนอกนี้ไม่ควรจะส่งผลต่อการจำแนกประเภทมากกว่าข้อมูลที่อยู่ด้านใน หรือมีน้ำหนักไม่มากกว่าข้อมูลที่อยู่ด้านใน



รูปที่ 3 ภาพตัวอย่างของการมีประโยชน์ไม่มากกว่า จะเห็นได้ว่าข้อมูลที่อยู่บนเส้นขอบของรูปทรงรอบนอกควรจะมีผลต่อการจำแนกประเภทไม่มากไปกว่าข้อมูลบนเส้นขอบของรูปทรงที่อยู่ด้านใน

คราวนี้ลองจินตนาการถึงเส้นตรงกลุ่มหนึ่งซึ่งผ่านจุดซึ่งแทนข้อมูลที่เราต้องการจำแนกประเภทแล้วตัดพื้นผิวของรูปทรงทั้งสองออกมาเป็นพื้นที่เล็กๆ โดยไม่สูญเสียลักษณะทั่วไปเราจะสามารถประมาณพื้นที่เหล่านี้ได้โดยใช้พื้นผิวส่วนหนึ่งของ ไฮเปอร์สเฟียร์ (hypersphere) ให้รัศมีของไฮเปอร์สเฟียร์ทั้งสองเป็น r_1 และ r_2 โดยที่ $r_1 \le r_2$ และพื้นที่ของพื้นผิวเล็กๆทั้งสองเป็น a_1 และ a_2 (ดูรูปที่ 3) จากข้อจำกัดของการมีประโยชน์ไม่มากกว่าและสมมุติฐานของการกระจายตัวแบบสม่ำเสมอของข้อมูล (uniform distribution) ในทุกๆคุณสมบัติ ซึ่งหมายความว่าค่าคาดหวังของจำนวนข้อมูลบนพื้นที่หนึ่งจะเป็นสัดส่วนโดยตรงกับขนาดของพื้นที่เหล่านั้นเราจะได้ว่า

$$f(r_1)a_1 \ge f(r_2)a_2$$

เนื่องจากขนาดของพื้นผิวของไฮเปอร์สเฟียร์ใดๆ ที่มีรัศมี r จะแปรผันตาม r^{n-1} เราจะได้ว่า

$$f(r_1)r_1^{\,n-1} \ge f(r_2)r_2^{\,n-1}$$

$$(r_1^{-t})r_1^{n-1} \geq (r_2^{-t})r_2^{n-1}$$

$$r_1^{n-1-t} \geq r_2^{n-1-t}$$

$$\left(\frac{r_1}{r_2}\right)^{n-1-t} \geq 1$$

$$\left(\frac{r_2}{r_1}\right)^{t-(n-1)} \geq \left(\frac{r_2}{r_1}\right)^{0}$$

$$t-(n-1) \geq 0$$

$$t \geq n-1$$

จากข้อจำกัดเพิ่มเติมนี้เราจะได้ฟังก์ชันใหม่คือ

$$f(x)=\frac{1}{x^{t}};t \geq n-1$$

จะสังเกตได้ว่าเครื่องหมาย $\geq$ นั้นเป็นผลมาจากการที่เรายอมผ่อนปรนให้ข้อมูลบนพื้นผิวทั้งสองอาจจะมีความสำคัญต่อผลลัพธ์เท่ากันได้ซึ่งไม่น่าจะเป็นจริง แต่เนื่องจากในการกำหนดรูปแบบทั่วไปของฟังก์ชันต่อจากนี้สมควรที่จะใช้ค่าขอบเขตล่างของ t ด้วย ดังนั้นจึงได้ทำการลดขอบเขตของข้อกำหนดในส่วนนี้เพื่อรวมค่า $t=n-1$ เข้าไปในการทดสอบด้วย

อย่างไรก็ตามหากต้องการที่จะพิจารณาเงื่อนไขของการมีประโยชน์ไม่มากกว่าอย่างแท้จริงแล้วเราจะได้ว่าข้อมูลทั้งหมดบนพื้นผิวของรูปทรงที่อยู่ด้านในไม่เพียงน่าจะมีความสำคัญมากกว่าข้อมูลทั้งหมดที่อยู่บนพื้นผิวของรูปทรงใดๆ ที่อยู่รอบนอกเท่านั้น แต่ควรจะมีความสำคัญสูงกว่าข้อมูลที่อยู่ด้านนอกเส้นขอบนั้นทั้งหมด แต่เงื่อนไขนี้จะไม่สามารถนำมาเปลี่ยนให้เป็นเป็นจริงได้เนื่องจากพื้นผิวของรูปทรงจะอยู่ใน $n-1$ มิติขณะที่พื้นที่ส่วนที่อยู่รอบนอกจะอยู่ใน n มิติ ดังนั้นเมื่อคำนวณค่าคาดหวังของอัตราส่วนของจำนวนข้อมูลที่คาดหวังว่าจะอยู่บนพื้นผิวต่อจำนวนข้อมูลที่อยู่ในพื้นที่รอบนอกจึงได้เป็น 0

เราสามารถหาทางออกได้โดยการกำหนดความหนาของพื้นที่ผิวของรูปทรงเป็นค่าคงที่ j ค่าหนึ่งซึ่งเมื่อรวมเข้ากับข้อกำหนดในเบื้องต้นและการสมมุติให้รูปทรงนี้เป็นไฮเปอร์สเฟียร์ที่มีรัศมี R ซึ่งจะไม่ทำให้สูญเสียลักษณะทั่วไป ให้ $A(R)$ เป็นพื้นที่ผิวของไฮเปอร์สเฟียร์ที่มีรัศมี R และ $V(R)$ เป็นปริมาตรของไฮเปอร์สเฟียร์ที่มีรัศมี R แล้วจะทำให้ได้ว่า

$$jA(R)f(R) \geq \int_{r=R}^{r=\infty} f(r)d(V(r))$$

$$jA(R)f(R) \geq \int_{R}^{\infty} f(r)A(r)dr$$

$$j(R^{n-1})f(R) \geq \int_R^{\infty} f(r)(r^{n-1})dr$$

ให้ $g(r) = r^{n-1} f(r)$

$$jg(R) \geq \int_R^{\infty} g(r)dr$$

เนื่องจากผลรวมจากข้อมูลทั้งหมดที่อยู่ไกลเกินไปจะไม่มีผลต่อค่าน้ำหนักนั่นก็คือ

$$\lim_{r\to\infty} r^{n-1} f(r) = 0$$

$$\lim_{r\to\infty} g(r) = 0$$

ดังนั้นเราจะได้ว่า

$$\frac{d}{dr}\left(\int_r^{\infty} g(r)dr\right) = \lim_{x\to\infty} g(x) - g(r) = -g(r)$$

จาก

$$jg(R) \geq \int_R^{\infty} g(r)dr$$

$$\left.\frac{g(r)}{\int_r^{\infty} g(r)dr}\right|_{r=R} \geq \frac{1}{j}$$

เนื่องจากค่าของ R สามารถเป็นค่าใดๆ ก็ได้โดยไม่มีผลต่ออสมการนี้ เราจึงสามารถทำการละไว้ได้

$$\frac{g(r)}{\int_r^{\infty} g(r)dr} \geq \frac{1}{j}$$

$$\frac{d}{dr}(-\ln(\int_r^{\infty} g(r)dr)) \geq \frac{1}{j}$$

$$-\ln(\int_r^{\infty} g(r)dr) \geq \int \frac{1}{j}\,dr$$

$$\ln(\int_r^{\infty} g(r)dr) \leq -\int \frac{1}{j}\,dr$$

$$\ln(\int_r^{\infty} g(r)dr) \leq -\frac{r}{j} + C$$

โดยที่ C เป็นค่าคงที่ของการอินทิเกรต

$$\ln(\int_r^\infty g(r)dr) \le -\frac{r}{j} + C$$

$$\int_r^\infty g(r)dr \le e^{-\frac{r}{j}+C}$$

ให้ $\int_r^\infty h(r)dr = e^{-\frac{r}{j}+C}$

$$\frac{d}{dr}(\int_r^\infty h(r)dr) = (\frac{d}{dr}e^{-\frac{r}{j}+C})$$

$$-h(r) = -\frac{1}{j}e^{-\frac{r}{j}+C}$$

$$h(r) = \frac{1}{j}e^{-\frac{r}{j}+C}$$

เราใช้ $h(r)$ เพื่อที่จะประมาณ $g(r)$ โดย

$$\int_r^\infty g(r)dr \le e^{-\frac{r}{j}+C}$$

$$\int_r^\infty g(r)dr \le \int_r^\infty h(r)dr$$

$$\frac{d}{dr}(\int_r^\infty g(r)dr) \le \frac{d}{dr}(\int_r^\infty h(r)dr)$$

$$-g(r) \le -h(r)$$

จึงได้ว่า

$$g(r) \ge h(r) = \frac{1}{j}e^{-\frac{r}{j}+C}$$

จากนั้นทำการเปลี่ยนตัวแปร r เป็น x และแทนค่าฟังก์ชัน $f(x)$

$$g(r) = r^{n-1}f(r)$$

$$f(x) = \frac{g(x)}{x^{n-1}} \ge (\frac{1}{x^{n-1}})\frac{1}{j}e^{-\frac{r}{j}+C}$$

นั่นก็คือเราจะได้รูปแบบของฟังก์ชันใหม่คือ

$$f(x) \geq \frac{1}{jx^{n-1}} e^{-\frac{x}{j}+C}$$

ทำการแทนค่าสัญลักษณ์โดยให้ $t = \frac{1}{j}$ และทำการตัดค่าสัมประสิทธิ์ที่คูณอยู่คือ $\frac{1}{j}e^{C}$ ออกเนื่องจากจะไม่มีผลต่อการจำแนกประเภท และจากการที่ $j > 0$ เนื่องจากความหนาของไฮเปอร์สเฟียร์จะต้องมากกว่า 0 เพื่อให้บริเวณที่อินทิเกรตมีจำนวนมิติเท่ากับบริเวณที่อยู่รอบนอก เราจะได้ว่า $t > 0$ ซึ่งก็คือเราจะสามารถเปลี่ยนรูปแบบของฟังก์ชันนี้ได้เป็น

$$f(x) \geq \frac{1}{x^{n-1}} e^{-tx}; t > 0$$

เนื่องจากความสัมพันธ์ของฟังก์ชันแบบมากกว่าหรือเท่ากับจะมีขอบเขตที่จะต้องพิจารณาค่อนข้างกว้าง เราจะทำการพิจารณากรณีที่เป็นกรณีเท่ากับเป็นพิเศษ ซึ่งก็คือ

$$f(x) = \frac{1}{x^{n-1}} e^{-tx}; t > 0$$

นอกจากนี้แล้วเราจะสังเกตเห็นข้อเท็จจริงได้ประการหนึ่ง นั่นก็คือในนิพจน์ e^{-tx} ซึ่ง $t > 0$ และ $x \geq 0$ เราจะได้ว่า $e^{-tx} \leq 1$

จาก $e^{-tx} \leq 1$ เราจะได้ว่า

$$\frac{1}{x^{n-1}} \geq \frac{1}{x^{n-1}} e^{-tx}$$

เราจะสังเกตได้ว่า ส่วนที่อยู่ด้านซ้ายของอสมการนี้เป็นกรณีเฉพาะของรูปแบบแรกนั่นคือ $f(x) = \frac{1}{x^{t}}; t \geq n-1$ เมื่อ $t = n-1$ และเมื่อพิจารณาค่าของอสมการข้างต้นในกรณีที่ $x = \lim_{z \to 0^{+}} z$ เราจะว่า ฟังก์ชัน $f(x) = \frac{1}{x^{n-1}}$ จะเป็นฟังก์ชันเดียวที่มีลักษณะสอดคล้องกับทั้งสมการของฟังก์ชันในรูปแบบแรก และอสมการของฟังก์ชันในรูปแบบที่สอง

ซึ่งหากลองมาใช้รูปแบบทั่วไปของฟังก์ชันที่ได้มาจากเงื่อนไขของการทนทานต่อการสเกลมาแทนลงไปในอสมการของเงื่อนไขนี้

$$j(R^{n-1}) f(R) \geq \int_{R}^{\infty} f(r)(r^{n-1}) dr$$

$$j(R^{n-1})(\frac{1}{R^{t}}) \geq \int_{R}^{\infty} (\frac{1}{r^{t}})(r^{n-1}) dr$$

$$j(R^{n-t-1}) \geq \int_{R}^{\infty} (r^{n-t-1}) dr$$

$$jR^{n-t-1} \geq \frac{1}{n-t}(\lim_{r\to\infty} r^{n-t} - R^{n-t})$$

$$jR^{n-t-1} \geq \frac{R^{n-t}}{t-n}$$

$$j(t-n) \geq R$$

เนื่องจาก R สามารถมีค่ามากได้ถึง ∞ ดังนั้นอสมการนี้จึงเป็นเท็จ

นั่นก็คือถึงแม้ว่าเราจะสามารถหาข้อจำกัดของ $f(x)$ โดยประมาณจากเงื่อนไขนี้ได้ แต่ข้อจำกัดนี้จะทำให้เกิดความขัดแย้งต่อข้อจำกัดของการทนทานต่อการสเกลซึ่งเป็นคุณสมบัติหนึ่งที่ต้องการ

นอกจากนี้หากเราทำการพิจารณาสมการของฟังก์ชันในรูปแบบที่สอง เราจะพบว่าสมการของฟังก์ชันรูปแบบนี้คือ $f(x) = \frac{1}{x^{n-1}} e^{-tx}; t > 0$ มีลักษณะที่คล้ายคลึงกับสมการของการแจกแจงความน่าจะเป็นแบบแกมมา (gamma probability distribution) [14] ซึ่งเป็นกรณีทั่วไปของการแจกแจงความน่าจะเป็นแบบเอกซโพเนนเชียล (exponential probability distribution) [14] และการแจกแจงความน่าจะเป็นแบบไคสแควร์ (chi-square probability distribution) [14] โดยสมการของการแจกแจงความน่าจะเป็นแบบแกมมาสามารถเขียนออกมาได้ว่า

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} e^{-x/\beta} x^{\alpha-1}; x > 0$$

$$= 0; x \leq 0$$

โดยที่ค่าปรับแต่งทั้งสองคือ α และ β จะมีค่ามากกว่า 0 และ $\Gamma(\alpha)$ คือฟังก์ชันแกมมา (gamma function) หรือฟังก์ชันทั่วไปของฟังก์ชันแฟกทอเรียล (generalized factorial function) [9,14,15,16,17,18] ซึ่งสามารถเขียนออกมาเป็นสมการได้ว่า

$$\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1} dx; \alpha > 0$$

หรือ $\Gamma(\alpha+1) = \int_0^{\infty} e^{-x} x^{\alpha} dx; \alpha > 0$

ซึ่งจากสมการนี้เองเราจะทำการอินทิเกรตทีละส่วน (integration by part) [19] โดยกำหนดให้

$u = e^{-x}$ และ $dv = x^{\alpha-1} dx$

ซึ่งจะทำให้ได้ว่า

$$du = -e^{-x}dx \text{ และ } v = \frac{x^{\alpha}}{\alpha}$$

จากนั้นทำการแทนค่าลงไปในการอินทิกรัล

$$\int udv = uv - \int vdu$$

$$\int_0^{\infty} e^{-x}x^{\alpha-1}dx = \left.\frac{e^{-x}x^{\alpha}}{\alpha}\right|_0^{\infty} - \int_0^{\infty}\frac{-e^{-x}x^{\alpha}}{\alpha}dx$$

$$\int_0^{\infty} e^{-x}x^{\alpha-1}dx = \frac{1}{\alpha}\int_0^{\infty} e^{-x}x^{\alpha}dx$$

$$\Gamma(\alpha) = \frac{1}{\alpha}\Gamma(\alpha+1)$$

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

ซึ่งเป็นความสัมพันธ์ในรูปแบบเดียวกับ

$$n! = n(n-1)!$$

และเมื่อทำโดยการอินทิเกรตเราจะได้ว่า

$$\Gamma(1) = \int_0^{\infty} e^{-x}x^{1-1}dx = \int_0^{\infty} e^{-x}dx = 1$$

ซึ่งเมื่อสรุปแล้วเราจะได้ว่า

$\Gamma(\alpha) = (\alpha-1)!$ เมื่อ α เป็นจำนวนเต็มบวก

และจากสมการของการแจกแจงความน่าจะเป็นแกมมานี้ หากทำการแทนค่าตัวแปรลงไปโดยให้ $\alpha = 1$ และพิจารณาเฉพาะเมื่อ $x > 0$ เราจะได้สมการ $f(x) = \frac{1}{\beta^{\alpha}}e^{-x/\beta}$ ซึ่งเป็นสมการของการแจกแจงความน่าจะเป็นแบบเอกซโพเนนเชียล และหากกำหนดให้ $\alpha = \frac{\upsilon}{2}$ และ $\beta = 2$ พิจารณาเฉพาะเมื่อ $x > 0$ เราก็จะได้สมการ $f(x) = \frac{1}{2^{\frac{\upsilon}{2}}\Gamma(\frac{\upsilon}{2})}e^{-\frac{x}{2}}x^{\frac{\upsilon}{2}-1}$ ซึ่งเป็นสมการของการแจกแจงความน่าจะเป็นแบบไคสแควร์

คราวนี้ลองพิจารณาสมการของการแจกแจงความน่าจะเป็นแบบแกมมาในฐานะที่เป็นฟังก์ชันถ่วงน้ำหนัก เมื่อทำการตัดค่าสัมประสิทธิ์ของฟังก์ชันและแทนค่าให้ $\beta = \frac{1}{t}$ เมื่อพิจารณาเฉพาะกรณีที่ $x > 0$ เราจะได้ว่า

$$f(x) = e^{-tx}x^{\alpha-1}; t > 0$$

ซึ่งจะต่างจากสมการของฟังก์ชันในรูปแบบที่ 2 คือ $f(x) = \frac{1}{x^{n-1}}e^{-tx}; t > 0$ เฉพาะส่วนที่เป็นเลขยกกำลังของ x โดยฟังก์ชันทั้งสองนี้จะเหมือนกันก็ต่อเมื่อ $\alpha = 2 - n$ แต่จากข้อกำหนดของการแจกแจงความน่าจะเป็นแบบแกมมาที่ว่า $\alpha > 0$ สมการนี้จึงจะเป็นจริงได้ก็ต่อเมื่อ $n = 1$ เท่านั้น

เมื่อพิจารณาต่อไป เราจะพบได้ว่าการกำหนดความหนาของพื้นที่ผิวของรูปทรงเป็นค่าคงที่ j ค่าหนึ่งจะมีข้อเสียประการหนึ่ง นั่นคือการที่จะไม่เกิดประโยชน์หากเราต้องการพิจารณาข้อมูลซึ่งน่าจะไม่มีประโยชน์ต่อการจำแนกประเภท ณ จุดพิกัดซึ่งมีระยะห่างจากจุดกึ่งกลางของรูปทรงไฮเปอร์สเฟียร์ที่กล่าวมาน้อยกว่าค่าคงที่ j ที่ได้กล่าวถึงไปแล้วในข้างต้น หรือกล่าวอีกนัยหนึ่ง เราจะไม่สามารถที่จะถือว่าข้อมูลที่มีความระยะห่างจากจุดซึ่งเป็นที่สนใจน้อยกว่าค่ารัศมี j ที่กำหนดไว้ไร้ประโยชน์ได้ เพื่อแก้ไขข้อจำกัดนี้เราสามารถที่จะใช้วิธีการกำหนดค่าคงที่ j หลังจากที่พิจารณาข้อมูลสอนทั้งหมดได้เช่นกัน

อย่างไรก็ตาม แทนที่จะใช้วิธีการพิจารณาข้อมูลสอนทั้งหมดก่อนที่จะหาค่าคงที่ j ที่เหมาะสมตามข้อกำหนดนี้ เราสามารถแก้ไขข้อกำหนดนี้ได้โดยกำหนดให้ความหนาของรูปทรงที่กำหนดมีการแปรเปลี่ยนไปตามระยะห่างจากจุดที่เป็นศูนย์กลางของไฮเปอร์สเฟียร์ นั่นคือเราจะใช้วิธีการกำหนดค่าคงที่ $\alpha \in [0,1]$ ขึ้นมา โดยถ้าระยะห่างระหว่างจุดศูนย์กลางของไฮเปอร์สเฟียร์กับของรอบนอกของรูปทรงเป็น x เราจะกำหนดให้ขอบในของรูปทรงนี้มีระยะห่างจากจุดศูนย์กลางของไฮเปอร์สเฟียร์เป็น αx นั่นก็คือเราจะได้ว่า $j = x - \alpha x = (1-\alpha)x$ และเนื่องจากช่วงที่เป็นไปได้ของค่า α จากการที่ $\alpha \in [0,1]$ เราจะได้ว่า $j \in [0, x]$ ซึ่งการที่ $j \le x$ นี้แสดงว่าเราสามารถใช้วิธีการกำหนดค่าอัตราส่วน α มาแก้ไขปัญหาดังกล่าวได้

กรณีที่ $\alpha = 0$ เราจะได้ว่าน้ำหนักโดยรวมของรูปทรงไฮเปอร์สเฟียร์ที่มีความหนาเท่ากับ j จะมีค่าไม่น้อยกว่าน้ำหนักโดยรวมของปริมาตรที่อยู่รอบนอกทั้งหมด จากการที่ค่าของ j สามารถหาได้จากสมการ $j = (1-\alpha)x = (1-0)x = x$ ซึ่งค่า x นั้นก็คือระยะห่างระหว่างจุดศูนย์กลางของไฮเปอร์สเฟียร์กับจุดที่สนใจใดๆ ซึ่งสามารถที่จะเป็นค่าใดๆก็ได้ที่อยู่ในช่วง $[0,\infty)$ นั่นก็คือเราจะได้ว่า j ซึ่งเป็นความหนาของไฮเปอร์สเฟียร์นั้นจะเป็นจำนวนจริงใดๆ ก็ได้ที่ไม่ติดลบ

จากการที่ $j \in [0,\infty)$ เรากำหนดให้ $j = \lim_{z \to 0^+} z$ หรือเป็นค่าที่ใกล้ 0 มากๆ เราก็จะยังได้ผลลัพธ์เช่นเดิมว่าน้ำหนักรวมของไฮเปอร์สเฟียร์นั้นจะยังคงไม่น้อยกว่าน้ำหนักของปริมาตรทั้งหมด แม้ว่าขอบเขตของไฮเปอร์สเฟียร์นั้นจะเข้าใกล้ 0 ก็ตาม นี่แสดงว่า ฟังก์ชันถ่วงน้ำหนักนี้จะมีลักษณะประการหนึ่งก็คือบริเวณที่เป็นศูนย์กลางของไฮเปอร์สเฟียร์หรือ $x = 0$ จะต้องมีลักษณะ

แบบเดียวกับฟังก์ชันเดลต้า (delta function) [16,17] ซึ่งมีค่าผลรวมเมื่ออินทิเกรตบริเวณที่ $x = 0$ ผ่านมากกว่าบริเวณที่เหลือทั้งหมดรวมกัน โดยฟังก์ชันเดลต้าหรือที่เรียกกันอีกอย่างว่าฟังก์ชันอิมพัลส์ซึ่งเราสามารถนิยามในทางคณิตศาสตร์ได้ดังนี้

$$\delta(x) = \infty; x = 0$$

$$\delta(x) = 0; x \neq 0$$

$$\int_{-\infty}^{\infty} \delta(x)dx = 1$$

หรือเป็นฟังก์ชันที่มีข้อกำหนดว่าเมื่ออินทิเกรตผ่านจุดที่ $x = 0$ จะมีค่าของการอินทิเกรตเปลี่ยนไปเท่ากับ 1 ทันที ขณะที่บริเวณอื่นของฟังก์ชันนี้ซึ่ง $x \neq 0$ จะมีค่าเป็น 0 ทั้งหมด ซึ่งนั่นก็คือเราจะได้ว่า

$$\int_{0}^{j} x^{n-1} f(x)dx \geq \int_{j}^{\infty} x^{n-1} f(x)dx \text{ เมื่อ } j = \lim_{z \to 0^+} z \quad \text{และ } n = \text{จำนวนมิติ}$$

โดยนิพจน์ x^{n-1} เป็นผลมาจากจำนวนมิติของขอบเขตของรูปทรงที่เป็นไฮเปอร์สเฟียร์

ซึ่งก็คือ

$x^{n-1} f(x) = \beta\delta(x) + \gamma$ ในช่วงที่ $x \in [0, j)$ สำหรับค่าคงที่ β ค่าหนึ่ง ซึ่ง $\beta \geq \int_{j}^{\infty} x^{n-1} f(x)dx$ และค่าคงที่ $\gamma = f(j)$ ซึ่งจะไม่มีผลต่อการอินทิเกรตเนื่องจาก $j = \lim_{z \to 0^+} z$ ทำให้พื้นที่ของการอินทิเกรตเล็กมาก

และเนื่องจากลักษณะของฟังก์ชันเดลต้านั้นเป็นฟังก์ชันไม่ต่อเนื่องเราจึงได้ว่า $\beta\delta(x) + \gamma$ ซึ่งเป็นผลจากการคูณและบวกฟังก์ชันเดลต้าด้วยค่าคงที่จะเป็นฟังก์ชันไม่ต่อเนื่องเช่นกัน ขณะเดียวกัน จากการที่ x^{n-1} เป็นฟังก์ชันต่อเนื่อง และสมการ $x^{n-1} f(x) = \beta\delta(x)$ เราจึงสรุปได้ว่า $f(x)$ จะต้องเป็นฟังก์ชันไม่ต่อเนื่อง เนื่องมาจากผลลัพธ์ที่ได้จากการคูณกันระหว่างฟังก์ชันต่อเนื่องกับฟังก์ชันต่อเนื่องจะต้องเป็นฟังก์ชันต่อเนื่องเท่านั้น

จากการวิเคราะห์ที่ผ่านมา ในกรณีที่ $\alpha = 0$ เราจะสามารถสรุปออกมาเป็นข้อจำกัดได้เพียง 2 ประการคือ

1. $f(x) = (\beta\delta(x) + \gamma) / x^{n-1}; x \in [0, \lim_{z \to 0^+} z)$
2. $f(x)$ เป็นฟังก์ชันไม่ต่อเนื่อง

ซึ่งจากข้อสรุปทั้งสองนี้ ข้อสรุปแรกเป็นเงื่อนไขที่มีประโยชน์น้อยมากเนื่องจากเป็นความสัมพันธ์ที่แสดงความเกี่ยวข้องกันระหว่างค่าของ $f(x)$ ในช่วงที่ x มีค่าน้อยมากๆ กับ

ผลที่ได้จากค่าโดยรวมของ $f(x)$ ในช่วงของ x ที่เหลือทั้งหมด โดยไม่แสดงถึงความสัมพันธ์กันระหว่างค่าของ x ในส่วนที่เหลือเมื่อนำมาเปรียบเทียบกัน และข้อสรุปที่สองซึ่งก็คือการที่ $f(x)$ เป็นฟังก์ชันไม่ต่อเนื่องนั้นจะขัดแย้งกับคุณสมบัติเบื้องต้นข้อหนึ่งที่ว่า $f(x)$ จะต้องเป็นฟังก์ชันต่อเนื่องชนิดไม่เพิ่มอย่างสิ้นเชิง ด้วยเหตุเหล่านี้เองเราจึงสรุปว่า $\alpha \neq 0$

กรณีที่ $\alpha = 1$ เราจะได้ว่า $j = (1-\alpha)x = (1-1)x = 0$ นั่นก็คือความหนาของรูปทรงจะมีค่าเท่ากับ 0 ซึ่งจะทำให้รูปทรงนี้จะกลายเป็นไฮเปอร์เพลนหรือระนาบที่มี $n-1$ มิติ ขณะที่บริเวณที่อยู่รอบนอกนั้นจะยังมีปริมาตรอยู่ใน n มิติ ดังนั้นน้ำหนักที่ได้จากไฮเปอร์เพลนนี้จึงไม่สามารถนำไปเทียบกับน้ำหนักทั้งหมดของบริเวณที่อยู่รอบนอกได้ แสดงว่า $\alpha \neq 1$

นั่นก็คือเราทำการเปลี่ยนช่วงของค่า α ให้แคบลงได้เป็น $\alpha \in (0,1)$

เช่นเดียวกับในการพิสูจน์ที่ได้ผ่านมาแล้ว เราทำการสมมุติให้รูปทรงนี้เป็นไฮเปอร์สเฟียร์ที่มีรัศมี R ซึ่งจะไม่ทำให้สูญเสียลักษณะทั่วไป ให้ $A(R)$ เป็นพื้นที่ผิวของไฮเปอร์สเฟียร์ที่มีรัศมี R และ $V(R)$ เป็นปริมาตรของไฮเปอร์สเฟียร์ที่มีรัศมี R แล้วจะทำให้ได้ว่า

$$\int_{r=\alpha R}^{r=R} f(r)d(V(r)) \geq \int_{r=R}^{r=\infty} f(r)d(V(r))$$

$$\int_{\alpha R}^{R} f(r)A(r)dr \geq \int_{R}^{\infty} f(r)A(r)dr$$

$$\int_{\alpha R}^{R} f(r)(r^{n-1})dr \geq \int_{R}^{\infty} f(r)(r^{n-1})dr$$

ให้ $g(r) = r^{n-1}f(r)$

$$\int_{\alpha R}^{R} g(r)dr \geq \int_{R}^{\infty} g(r)dr$$

ให้ $G(s) = \int g(r)dr\Big|_{r=s}$

$$G(R) - G(\alpha R) \geq \lim_{r\to\infty} G(r) - G(R)$$

ทำการกำหนดค่าคงที่ของการอินทิเกรตของ $\int g(r)dr$ โดยให้ $\lim_{r\to\infty} G(r) = 0$

$$G(R) - G(\alpha R) \geq -G(R)$$

$$2G(R) \geq G(\alpha R)$$

จากการที่เราจะได้ว่า $f(r) \geq 0$ ทำให้ $g(r) = r^{n-1} f(r) \geq 0$ และเนื่องมาจาก เรากำหนดให้ $G(r) = \int g(r)dr$ ทำให้ $g(r) = \frac{d}{dr} G(r)$ นั่นก็คือ $g(r)$ เป็นฟังก์ชันความชันของ $G(r)$ ด้วยเหตุนี้เองเมื่อ $g(r) = r^{n-1} f(r) \geq 0$ จึงทำให้ $G(r)$ จึงเป็นฟังก์ชันไม่ลด

เนื่องมาจากการที่ $G(r)$ เป็นฟังก์ชันไม่ลดนี้เองทำให้เราสามารถที่จะสรุปได้ว่า $\lim_{r\to\infty} G(r) \geq G(R) \geq G(\alpha R)$ และด้วยการที่เรากำหนดค่าคงที่ในการอินทิเกรตของ $G(r)$ ไปโดยทำให้ $\lim_{r\to\infty} G(r) = 0$ นี้เองเราจึงได้ว่า $G(\alpha R) \leq G(R) \leq 0$ ซึ่งก็คือ $G(R)$ และ $G(\alpha R)$ จะต้องไม่เป็นจำนวนจริงบวก

จาก $2G(R) \geq G(\alpha R)$ เราจึงได้ว่า

$$\frac{2G(R)}{G(\alpha R)} \leq 1$$

$$\frac{G(\alpha R)}{G(R)} \geq 2$$

เพื่อความง่ายเราจะพิจารณา เราจะทำการพิจารณากรณีเฉพาะซึ่ง $\frac{G(\alpha R)}{G(R)} = m \geq 2$ นั่นคือ กรณีที่อัตราส่วนระหว่างค่าของฟังก์ชัน $G(r)$ เป็นค่าคงที่เนื่องจากการที่เป็นรูปแบบที่ง่ายที่สุดตามหลักใบมีดของออคแคม (Occam's razor) [1] ที่ว่ารูปแบบของการอธิบายลักษณะของระบบหรือสมมุติฐานที่มีความซับซ้อนน้อยจะมีความน่าจะเป็นที่จะถูกต้องสูงกว่ารูปแบบหรือสมมุติฐานที่มีความซับซ้อนมาก

จากสมการ $\frac{G(\alpha R)}{G(R)} = m$ สำหรับค่า R ใดๆ เราจะสามารถเห็นได้อย่างชัดเจนว่า $G(r)$ จะต้องอยู่ในรูปแบบ $G(r) = cr^{d}$ เมื่อ c และ d เป็นค่าคงที่

อีกประการหนึ่ง จากการที่ $G(\alpha R) \leq 0$ และ $G(R) \leq 0$ ซึ่งเป็นผลจากการกำหนดค่าคงที่ของการอินทิเกรตของ $G(r)$ ด้วยเงื่อนไขเดียวกันเราจึงสามารถที่จะสรุปได้ว่า $G(r) \leq 0$ เมื่อ $r \geq 0$ นั่นก็คือ เราจะได้ว่าค่าสัมประสิทธิ์ c ซึ่งเป็นค่าคงที่ซึ่งถูกนำไปคูณใน $G(r)$ จะต้องเป็นค่าติดลบ เพื่อความง่ายของสมการเราจึงจะเปลี่ยนรูปแบบของ $G(r)$ เป็น $G(r) = -cr^{d}$ เมื่อ $c \geq 0$

จากสมการ $\frac{G(\alpha R)}{G(R)} = m$

$$\frac{-c(\alpha R)^{d}}{-c(R)^{d}} = m$$

$$\alpha^{d} = m$$

$$d = \log_{\alpha} m$$

$$G(r) = -cr^{\log_{\alpha} m}$$

และจาก $G(r) = \int g(r)dr$

$$g(r) = \frac{d}{dr}G(r)$$

$$= \frac{d}{dr}(-cr^{\log_{\alpha} m})$$

$$= \frac{d}{dr}(-c\exp((\ln r)(\log_{\alpha} m)))$$

$$= (-cr^{\log_{\alpha} m})\frac{d}{dr}((\ln r)(\log_{\alpha} m))$$

$$= (-cr^{\log_{\alpha} m})(\log_{\alpha} m)(\frac{1}{r})$$

$$g(r) = (-c\log_{\alpha} m)r^{\log_{\alpha} m-1}$$

สุดท้าย จากสมการ $g(r) = r^{n-1}f(r)$

$$f(r) = \frac{g(r)}{r^{n-1}}$$

$$f(r) = \frac{(-c\log_{\alpha} m)(r^{\log_{\alpha} m-1})}{r^{n-1}}$$

$$f(r) = (-c\log_{\alpha} m)r^{\log_{\alpha} m-n}$$

เนื่องจาก $m \geq 2$

$$\ln m \geq \ln 2 > \ln 1 = 0$$

และจากการที่ $\alpha \in (0,1)$

$$\frac{1}{\alpha} \in (1,\infty)$$

$$\ln\alpha = -\ln(\frac{1}{\alpha}) \in (-\infty,0)$$

หรือ $\ln\alpha < 0$

นั่นก็คือ

$$\log_{\alpha} m = \frac{\ln m}{\ln\alpha} \leq 0$$

แสดงว่า สัมประสิทธิ์ $-c\log_{\alpha} m$ ของ $f(r)$ มีค่าไม่น้อยกว่า 0

ทำการจัดรูปแบบของสมการใหม่โดยการตัดสัมประสิทธิ์ของ $f(r)$ และแทนค่า $\log_{\alpha} m$ โดยให้ $w = -\log_{\alpha} m$ แล้วแทนที่ตัวแปร r ด้วย x หลังจากนั้นเราจะได้ว่า

$$f(x) = \frac{1}{x^{n+w}}; w \geq 0$$

ซึ่งเมื่อทำการแทนที่ $n+w$ ด้วย t แล้วเราจะสามารถเปลี่ยนรูปของสมการได้เป็น

$$f(x) = \frac{1}{x^{t}}; t \geq n$$

ซึ่งจะเห็นได้ว่านี่เป็นกรณีเฉพาะของรูปแบบ $f(x) = \frac{1}{x^{t}}; t \geq n-1$ ซึ่งได้มาจากการวิเคราะห์เมื่อเปรียบเทียบกันเฉพาะระหว่างข้อมูลที่อยู่บนขอบเขตของไฮเปอร์สเฟียร์ที่มีขนาดไม่เท่ากัน

และหากเราทำการแทนค่า m ด้วย 2 ซึ่งเป็นค่าขอบเขตของ m และให้ $\alpha = 0.5$ หรือพิจารณาบริเวณของไฮเปอร์สเฟียร์โดยที่ให้ขอบเขตข้างในมีรัศมีครึ่งหนึ่งของขอบเขตข้างนอกเราจะได้ว่า

$$f(x) = \frac{1}{x^{n-\log_{0.5} 2}}$$

$$= \frac{1}{x^{n+\log_{2} 2}}$$

$$f(x) = \frac{1}{x^{n+1}}$$

แต่หากเราใช้วิธีการกำหนดค่า α ตามขนาดของบริเวณที่แบ่งโดยให้ β เป็นอัตราส่วนของขนาดของบริเวณที่อยู่ภายในขอบเขตด้านในของไฮเปอร์สเฟียร์ต่อขนาดของบริเวณที่อยู่ภายในขอบเขตด้านนอกของไฮเปอร์สเฟียร์ เราจะได้ว่า

$$\alpha^{n} = \beta$$

$$\alpha = \sqrt[n]{\beta}$$

ซึ่งเมื่อแทนค่าเข้าไปในฟังก์ชัน $f(x)$ เราจะได้เป็น

$$f(x) = \frac{1}{x^{n-\log_{\alpha} m}}$$

$$= (x^{n-(\frac{\ln m}{\ln \alpha})})^{-1}$$

$$= (x^{n-n(\frac{\ln m}{\ln \beta})})^{-1}$$

$$f(x) = \frac{1}{x^{n(1-\ln_{\beta} m)}}$$

ซึ่งเมื่อทำการเปลี่ยนรูปโดยให้ $t = 1 - \ln_{\beta} m$ เราจะได้ว่า

$$f(x) = \frac{1}{x^{tn}}; t \geq 1$$

ซึ่งหากกำหนดให้ $m = 2$ และ $\beta = 0.5$ เราจะได้ว่า

$$f(x) = \frac{1}{x^{n(1-\ln_{0.5} 2)}}$$

$$= \frac{1}{x^{n(1+\ln_{2} 2)}}$$

$$= \frac{1}{x^{n(1+1)}}$$

$$f(x) = \frac{1}{x^{2n}}$$

สรุปแล้ว จากกระบวนการทั้งหมดที่ผ่านมา ในขณะนี้เราได้ข้อกำหนดของฟังก์ชันถ่วงน้ำหนักมา 3 รูปแบบได้แก่

1. $f(x) = \frac{1}{x^{t}}; t \geq n-1$
2. $f(x) = \frac{1}{x^{n-1}} e^{-tx}; t > 0$
3. $f(x) = \frac{1}{x^{tn}}; t \geq 1$

โดยข้อกำหนดของฟังก์ชันรูปแบบแรกมาจากเงื่อนไขของการมีความสำคัญไม่มากกว่าของข้อมูลที่อยู่ในวงรอบนอกซึ่งให้ผลที่คล้ายคลึงกับรูปแบบที่จากเงื่อนไขของการมีความสำคัญไม่มากกว่าของข้อมูลที่อยู่รอบนอกทั้งหมดเมื่อพิจารณาเทียบกับข้อมูลภายในโดยการกำหนดอัตราส่วนของระยะทาง ส่วนข้อกำหนดของฟังก์ชันที่มีรูปแบบที่สองมาจากเงื่อนไขของการมีความสำคัญไม่มากกว่าของข้อมูลที่อยู่รอบนอกทั้งหมดโดยการกำหนดค่าของระยะทางที่จะใช้ในการพิจารณา ซึ่งรูปแบบของฟังก์ชันประเภทที่สองนี้จะขาดคุณสมบัติในการทนทานต่อการสเกลซึ่งเป็นคุณสมบัติที่ดีอย่างหนึ่ง และฟังก์ชันในรูปแบบที่สามซึ่งเป็นผลจากเงื่อนไขของการมีความสำคัญไม่มากกว่าของข้อมูลที่อยู่รอบนอกทั้งหมดเมื่อพิจารณาเทียบกับข้อมูลภายในโดยการกำหนดอัตราส่วนของบริเวณที่ครอบคลุม

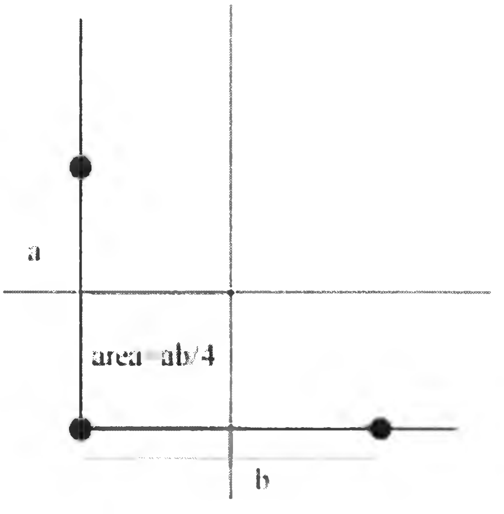
3.4 การประมาณการกระจายตัวแบบสม่ำเสมอของข้อมูล

คราวนี้ลองพิจารณาอีกสถานการณ์หนึ่ง โดยอาศัยสมมุติฐานที่ว่าข้อมูลทั้งหมดถูกสร้างขึ้นมาโดยการสุ่มจากคลัสเตอร์ (cluster) หลายๆ กลุ่มซึ่งเป็นสิ่งที่แสดงถึงลักษณะที่แท้จริงของข้อมูล ลองพิจารณาข้อมูลสองตัวที่มีความแตกต่างจากข้อมูลที่เราต้องการจำแนกประเภทเท่ากันเมื่อพิจารณาตามระยะทางระหว่างจุด ข้อมูลตัวหนึ่งเป็นของคลัสเตอร์ขนาดใหญ่ซึ่งหมายถึงคลัสเตอร์ที่มีความแปรปรวนของข้อมูลสูงในการสร้างข้อมูลซึ่งทำให้ข้อมูลที่ถูกสร้างออกมาน่าจะมีความแตกต่างกันได้มาก ขณะที่ข้อมูลอีกตัวหนึ่งมาจากคลัสเตอร์ที่มีขนาดเล็กซึ่งมีข้อมูลถูกสร้างออกมาจากคลัสเตอร์นี้ในปริมาณที่เท่ากันกับคลัสเตอร์ที่มีขนาดใหญ่กว่า ในกรณีเช่นนี้ข้อมูลทั้งสองนี้ควรจะมีผลต่อผลลัพธ์ของการจำแนกประเภทไม่เท่ากัน เนื่องมาจากการที่ข้อมูลในคลัสเตอร์ที่มีขนาดใหญ่จะมีผลต่อการทำนายข้อมูลที่อยู่รอบๆข้อมูลนั้นเองมากกว่า นั่นก็คือเซลล์หรือช่องของแผนภาพโวโรนอย ที่มีข้อมูลนั้นอยู่จะมีขนาดใหญ่กว่าของข้อมูลอีกอันหนึ่งและข้อมูลใดๆในเซลล์นี้ก็จะมีผลต่อผลลัพธ์ของการจำแนกประเภท เพราะการกระจายตัวของข้อมูลในอุดมคติสำหรับอัลกอริทึมเนียเรสท์เนเบอร์ก็คือการกระจายตัวแบบสม่ำเสมอเนื่องมาจากธรรมชาติของอัลกอริทึมนี้ซึ่งจะประมาณผลลัพธ์จากข้อมูลที่มีลักษณะคล้ายคลึงกัน ด้วยเหตุนี้จึงเป็นการสมควรที่เราจึงสมควรที่จะเอาเรื่องความสม่ำเสมอในการกระจายตัวของข้อมูลมาใช้ประโยชน์ในการจำแนกประเภทเช่นกัน โดยเราสามารถที่จะทำให้ผลลัพธ์ของการจำแนกประเภทจากการกระจายตัวของข้อมูลของเราเข้าใกล้การกระจายตัวแบบสม่ำเสมอหรือการแจกแจงความน่าจะเป็นแบบสม่ำเสมอ (uniform probability distribution) [14] ได้โดยการการเพิ่มน้ำหนักสำหรับข้อมูลซึ่งมีเซลล์ในแผนภาพโวโรนอยขนาดใหญ่

เราสามารถหาขนาดของเซลล์ในแผนภาพโวโรนอยที่มีขนาดใหญ่ได้แต่กระบวนการที่ใช้จะค่อนข้างซับซ้อนและใช้เวลาค่อนข้างมาก ในที่นี้เราจึงจะใช้วิธีการประมาณ และเนื่องจากในการใช้ค่าที่ได้ในการเพิ่มน้ำหนักของข้อมูลนั้นจะเหมือนกับการสร้างฟังก์ชัน $h(d)$ เป็นค่าปัจจัยสำหรับคูณกับ $f(x)$ เมื่อคำนวณน้ำหนักของข้อมูล d ในการจำแนกประเภทเพื่อทดแทนผลลัพธ์ที่มาจากขนาดของเซลล์ เพราะฉะนั้นหากจะนำวิธีนี้ไปใช้กับฟังก์ชันในรูปแบบที่ 1 หรือ 3 ซึ่งมีคุณสมบัติการทนทานต่อการสเกลแล้วเราจึงต้องระวังด้วยว่าฟังก์ชันที่เป็นผลลัพธ์จากการคูณกันของทั้งสองฟังก์ชันนี้จะยังคงมีคุณสมบัติการทนทานต่อการสเกลด้วยหากยังต้องการคงคุณสมบัตินี้ไว้

ให้ $h(d)$ เป็นค่าปัจจัยสำหรับคูณกับ $f(x)$ เมื่อคำนวณน้ำหนักของข้อมูล d ในการจำแนกประเภทเพื่อทดแทนผลลัพธ์ที่มาจากขนาดของเซลล์ วิธีการที่เราเลือกใช้ในการคำนวณ $h(d)$ นั้นคล้ายคลึงกับอัลกอริทึมเนียเรสท์เนเบอร์ดั้งเดิมซึ่งจะกำหนดจำนวนของข้อมูล m ตัว ซึ่งใกล้เคียงกับข้อมูลที่เราต้องการคำนวณค่า $h(d)$ ที่สุดจากระยะทางเมื่อถือว่าข้อมูลเป็นจุด เป็นที่ชัดเจนว่าขนาดของเซลล์สามารถที่จะประมาณได้จากการยกกำลังระยะทางระหว่างข้อมูลด้วยจำนวน

คุณสมบัติของข้อมูลซึ่งจะถูกเปลี่ยนเป็นจำนวนมิติในแผนภาพโวโรนอย วิธีการในอุดมคติที่จะรวมข้อมูลจากหลายๆจุดก็คือการใช้ค่าเฉลี่ยเรขาคณิตของระยะทางซึ่งสามารถแสดงให้เห็นได้โดยตัวอย่างต่อไปนี้ นั่นก็คือหากเราต้องการคำนวณขนาดของสี่เหลี่ยมผืนผ้าที่เกิดจากการตัดกันของเส้นที่แบ่งเซลล์ของแผนภาพโวโรนอยระหว่างจุดที่สนใจซึ่งอยู่ที่จุดตัดของแกนกับจุดที่อยู่ห่างออกไป a และ b ในแนวตั้งฉากเราจะได้ว่าพื้นที่ของรูปสี่เหลี่ยมผืนผ้าจะเป็น $ab/4$ (ดูรูปที่ 4) ซึ่งจะเป็นค่าที่แปรผันตามเป็นค่าเฉลี่ยเรขาคณิตของระยะห่างระหว่างจุดที่สนใจกับจุดทั้งสองยกกำลังด้วยจำนวนมิติ โดยที่จุดที่สนใจซึ่งอยู่ที่จุดตัดของแกนของรูปนี้ก็เป็นเหมือนข้อมูลที่เราต้องการหาค่า $h(d)$ ขณะที่จุดอีกสองจุดก็เป็นเหมือนกับข้อมูลที่อยู่ใกล้กับจุดนี้ โดยไม่สูญเสียลักษณะทั่วไปเราจะสามารถนำวิธีนี้ไปใช้ในการพิสูจน์กับรูปร่างต่างๆ ในมิติที่มากกว่านี้ได้เช่นกัน



รูปที่ 4 ภาพตัวอย่างสนับสนุนแนวคิดในการใช้ค่าเฉลี่ยเรขาคณิต จุดทั้ง 3 แสดงถึงตำแหน่งของข้อมูลสอน 3 ตัว โดยตัวที่อยู่ที่จุดตัดของแกนเป็นข้อมูลที่ต้องการหาขนาดของเซลล์ของแผนภาพโวโรนอย และเส้นตรงอีก 2 เส้นในรูปเป็นเส้นแบ่งขอบเขตของเซลล์ที่เกิดจากการพิจารณาข้อมูลที่สนใจกับข้อมูลโดยรอบทั้งสอง

คราวนี้เราจะทำการกำหนดค่า $h(d)$ ต่างๆ ที่จะใช้ในการทดลองเมื่อใช้ค่า $f(x)=\frac{1}{x^n}$ ซึ่งเป็นรูปแบบที่ง่ายที่สุดและให้ความแม่นยำใกล้เคียงกับ $f(x)=\frac{1}{x^{n-1}}$ ซึ่งให้ผลลัพธ์ที่ดีที่สุดในการทดลองที่ผ่านมา ให้ d_{ij} = ระยะทางระหว่างข้อมูลที่ใกล้กับ d_i เป็นอันดับที่ j กับ d_i (ไม่รวม d_i) และ m = จำนวนของข้อมูลที่จะใช้ในการคำนวณ $h(d)$

$$h_1(d_i)=\left(\prod_{j=1}^{m} d_{ij}\right)^{\frac{n}{m}} = (\text{ ค่าเฉลี่ยเรขาคณิตของระยะทาง })^n$$

$$h_2(d_i) = \frac{\sum_{j=1}^{m} d_{ij}^{\,n}}{m} = \text{ค่าเฉลี่ยเลขคณิตของ (ระยะทาง}^{n}\text{)}$$

$$h_3(d_i) = \left(\frac{\sum_{j=1}^{m} d_{ij}}{m} \right)^{n} = \text{(ค่าเฉลี่ยเลขคณิตของระยะทาง)}^{n}$$

ทฤษฎีบทที่ 2: ฟังก์ชันถ่วงน้ำหนัก $f(x)$ เมื่อคูณด้วย $h_1(d)$ $h_2(d)$ หรือ $h_3(d)$ จะให้ผลลัพธ์เป็นฟังก์ชันถ่วงน้ำหนักใหม่ที่ยังคงมีคุณสมบัติในการทนทานต่อการสเกลอยู่

พิสูจน์: เริ่มแรกเราจะทำการพิสูจน์ว่าทั้ง $h_1(d)$, $h_2(d)$ และ $h_3(d)$ ต่างก็ทนทานต่อการสเกล

$$h_1(ad) = \left(\prod_{j=1}^{m} ad_{ij}\right)^{\frac{n}{m}}$$

$$= \left(a^m \prod_{j=1}^{m} d_{ij}\right)^{\frac{n}{m}}$$

$$= a^n \left(\prod_{j=1}^{m} d_{ij}\right)^{\frac{n}{m}}$$

$$= a^n h_1(d)$$

$$h_2(d) = \frac{\sum_{j=1}^{m} (ad_{ij})^n}{m}$$

$$= \frac{\sum_{j=1}^{m} a^n d_{ij}^{\,n}}{m}$$

$$= a^n \frac{\sum_{j=1}^{m} d_{ij}^{\,n}}{m}$$

$$= a^n h_2(d)$$

$$h_3(ad) = \left(\frac{\sum_{j=1}^{m} ad_{ij}}{m} \right)^{n}$$

$$= \left(a \frac{\sum_{j=1}^{m} d_{ij}}{m} \right)^n$$

$$= a^n \left(\frac{\sum_{j=1}^{m} d_{ij}}{m} \right)^n$$

$$= a^n h_3(d)$$

ในขั้นตอนต่อไป เราจะพิสูจน์ว่าผลลัพธ์ของการคูณกันของฟังก์ชันใดๆ ที่ทนต่อการสเกล จะเป็นฟังก์ชันที่ทนต่อการสเกล

ให้ $s(x)$ และ $t(y)$ เป็นฟังก์ชันที่ทนต่อการสเกลซึ่งสำหรับอัตราส่วนของการสเกล a จะให้ผลลัพธ์ที่มีการสเกลเป็น α และ β

$$s(ax)t(ay) = (\alpha s(x))(\beta t(y))$$

$$= (\alpha\beta)(s(x)t(y))$$

โดยไม่สูญเสียลักษณะทั่วไป เราสามารถที่จะขยายผลจากการพิสูจน์นี้ไปใช้กับการคูณของฟังก์ชันจำนวนเท่าใดก็ได้

จากทั้งหมดนี้ เราจึงสามารถที่จะสรุปได้แล้วว่าเป็นการชัดเจนแล้วว่าฟังก์ชันที่เป็นผลลัพธ์จะทนต่อการสเกล □

นอกจากนี้ หากเราพิจารณาสมการ $f(x) = \frac{1}{x^{n-1}} e^{-tx}; t > 0$ ซึ่งเป็นฟังก์ชันถ่วงน้ำหนักรูปแบบที่สองที่ได้นำเสนอไป เนื่องจากสมการในรูปแบบนี้ไม่ทนต่อการเปลี่ยนสเกลของข้อมูล ดังนั้นเราจึงจะไม่สามารถหาค่าคงที่ t ที่เหมาะสมที่สุดได้ เนื่องมาจากการที่ค่า t นี้จะเปลี่ยนไปตามสเกลของคุณสมบัติ สังเกตได้ว่าต้นเหตุของการที่สมการนี้ไม่ทนต่อการเปลี่ยนสเกลนั้นเป็นเหตุมาจากนิพจน์ e^{-tx} แต่ว่าเราก็สามารถที่จะทำให้สมการนี้ทนทานต่อการสเกลได้โดยการกำหนดให้ $t \propto \frac{1}{\alpha}$ เมื่อให้ α เป็นอัตราส่วนในการเปลี่ยนสเกลของค่าในทุกมิติ เราสามารถสังเกตได้ว่าเราสามารถประมาณค่า α ได้โดยอาศัย $h(d)$ ซึ่งเป็นฟังก์ชันที่เราใช้ในการประมาณขอบเขตโดยรอบของข้อมูลสอนแต่ละตัว เพราะว่า $h(d) \propto \alpha^n$ โดยการอาศัยค่าที่ได้จากการเฉลี่ยกันของ $h(d)$ ของข้อมูลสอนทุกตัว ดังนั้นเราจึงสามารถที่จะเปลี่ยนรูปแบบของสมการแบบที่สองได้เป็น $f(x) = \frac{1}{x^{n-1}} \exp(-cx / \overline{\sqrt[n]{h(d)}}); c > 0$ เมื่อ c เป็นค่าคงที่ค่าหนึ่ง และ $\overline{\sqrt[n]{h(d)}}$ เป็นค่าเฉลี่ยของ

รากที่ n ของค่า $h(d)$ ซึ่งใช้ในการประมาณอัตราส่วนในการเปลี่ยนสเกล α ตามที่ได้กล่าวถึงไปแล้ว

จะสังเกตได้ว่าจากสมการที่ผ่านมานั้น เราสามารถนำค่าจากฟังก์ชัน $h(d)$ มาคูณเข้าไปเพื่อประมาณการกระจายตัวแบบสม่ำเสมอได้เช่นเดียวกับสมการในรูปแบบที่ 1 และ 3 แต่ว่าสมการนี้จะมีข้อแตกต่างอย่างหนึ่งก็คือจะมีนิพจน์ $\sqrt[n]{h(d)}$ ซึ่งเราใช้ในการประมาณอัตราส่วนของเปลี่ยนการสเกลอยู่ ด้วยเหตุนี้เองเราจึงสามารถที่จะทำการแทนค่า $h(d)$ ลงไปเพื่อที่จะประมาณการกระจายตัวแบบสม่ำเสมอของข้อมูลได้เลย ซึ่งหลังจากทำการแทนค่าแล้ว เราก็จะได้สมการผลลัพธ์ $f(x) = \frac{1}{x^{n-1}} \exp(-cx / \sqrt[n]{h(d)}); c > 0$

นั่นก็คือจากฟังก์ชันในรูปแบบที่ 2 เราจะได้รูปแบบที่แยกย่อยออกมาอีก 2 รูปแบบจากการกำหนดค่าคงที่ k เพื่อแก้ไขให้ฟังก์ชันนี้มีคุณสมบัติในการทนทานต่อการเปลี่ยนสเกล ได้แก่

1. $f(x) = \frac{1}{x^{n-1}} \exp(-cx / \sqrt[n]{\overline{h(d)}}); c > 0$
2. $f(x) = \frac{1}{x^{n-1}} \exp(-cx / \sqrt[n]{h(d)}); c > 0$