บทที่ 3

การจำแนกข้อมูลจากดาวเทียมและทฤษฎีสถิติที่เกี่ยวข้อง

3.1 วิธีการจำแนกประเภทข้อมูลจากดาวเทียม

การนำข้อมูลจากดาวเทียมสำรวจทรัพยากรมาใช้ในการศึกษาและวิเคราะห์นั้น มีขบวนการวิเคราะห์เพื่อให้ได้ตามวัตถุประสงค์คือ การจำแนกประเภทข้อมูล (Classification) ว่าวัตถุหรือสสารใดๆ บนพื้นโลกนั้นเป็นอะไร มีจำนวนหรือปริมาณเท่าใด ซึ่งแบ่งออกเป็น 2 วิธีตามลักษณะของการวิเคราะห์ ดังนี้

1. การแปลด้วยสายตา (Visual interpretation) เป็นการศึกษาและวิเคราะห์ข้อมูลจากดาวเทียม โดยการนำภาพถ่ายดาวเทียมมาแปลด้วยสายตาเพื่อแยกประเภทข้อมูล จากการพิจารณาระดับความเข้มของสี (Tone) ขนาด (Size) รูปร่าง (Shape) ความหยาบละเอียด (Texture) ตลอดจนสิ่งแวดล้อมข้างเคียง เป็นต้น ประกอบกับการสำรวจภาคพื้นดิน (Ground Truth) ก็สามารถที่จะแยกหรือจำแนกประเภทข้อมูลในแหล่งหรืออาณาบริเวณที่ต้องการว่าประกอบด้วยอะไรบ้าง สำหรับพื้นที่ของข้อมูลแต่ละประเภทวัดได้ด้วยเครื่องมือที่เรียกว่า Planimeter

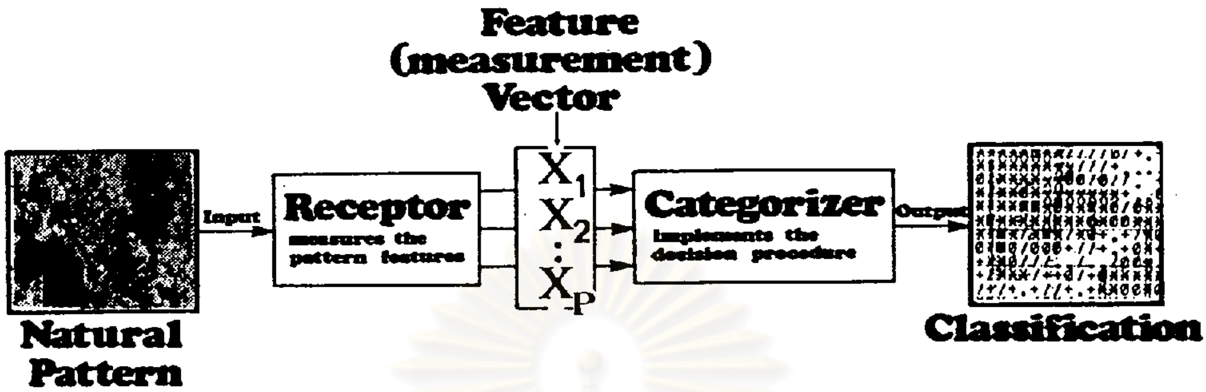
2. การวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ เป็นการศึกษาและวิเคราะห์ข้อมูลจากดาวเทียมโดยการนำเทปบันทึกข้อมูลจากดาวเทียม (CCT) มาทำการวิเคราะห์ด้วยเครื่องคอมพิวเตอร์ ซึ่งมีวิธีการวิเคราะห์ 2 วิธีการ กล่าวคือ (7 : 177)

2.1 Supervised analysis เป็นการวิเคราะห์โดยที่ผู้วิจัยจะต้องกำหนดพื้นที่ตัวอย่าง (Training area) โดยการสั่งการให้เครื่องคอมพิวเตอร์ได้รู้หรือรับทราบ ว่าพื้นที่ตัวอย่างจะต้องมีกี่ประเภทข้อมูล (Class) คอมพิวเตอร์จะอ่านค่าระดับสีเทา (Greytone level) ของข้อมูลแต่ละประเภทในพื้นที่ตัวอย่างมาคำนวณหาค่าสถิติ เช่น ค่าเฉลี่ย (mean) ค่าความแปรปรวนร่วม (covariance) ค่าสหสัมพันธ์ (correlation) เป็นต้น แล้วใช้ค่าสถิติต่างๆ เหล่านี้ไปแทนค่าในฟังชั่นทางสถิติที่กำหนด เพื่อใช้ใน

การจำแนกประเภทข้อมูลของอาณาบริเวณทั้งหมดที่ต้องการได้ ทั้งนี้พื้นที่ตัวอย่างกำหนดได้จากการตรวจสอบภาคพื้นดินแล้วว่า เป็นข้อมูลประเภทใดหรือเป็นวัตถุชนิดใด ซึ่งพื้นที่ตัวอย่างนี้จะต้องมีความเหมาะสมสามารถใช้เป็นตัวแทนของข้อมูลประเภทนั้นๆ ทั้งหมดอาณาบริเวณที่ศึกษา สำหรับการศึกษาครั้งนี้กำหนดเลือกใช้วิธีดังกล่าวซึ่งจะอธิบายในรายละเอียดต่อไป

2.2 Unsupervised analysis เป็นการวิเคราะห์โดยที่ไม่จำเป็นต้องกำหนดพื้นที่ตัวอย่างหรือสั่งการให้เครื่องคอมพิวเตอร์ได้รู้หรือรับทราบข้อมูลที่ต้องการจำแนกประเภทล่วงหน้าก่อนที่จะปฏิบัติงานเลย คอมพิวเตอร์จะทำการแยกประเภทข้อมูลทั้งหมดในอาณาบริเวณที่ศึกษาว่าควรจะมีกี่ประเภท ในลักษณะที่คล้ายๆ กับการกำหนดพื้นที่ตัวอย่างโดยอัตโนมัติ แล้วทำการจำแนกประเภทข้อมูลเพื่อให้ได้ผลลัพธ์ขั้นสุดท้าย อนึ่งการวิเคราะห์ด้วยวิธีดังกล่าวนี้ ใช้ได้กับโปรแกรมสำเร็จรูปบางโปรแกรมเท่านั้น ซึ่งโปรแกรมสำเร็จรูป MOA - RECOGX ไม่สามารถกระทำได้

จากรายละเอียดที่กล่าวข้างต้นนี้ เมื่อนำมาอธิบายในลักษณะรูปแบบจำลอง (Model) ของวิธีการจำแนกประเภทข้อมูลได้ดังนี้ จะเห็นว่าวัตถุบนพื้นผิวโลกจะถูกบันทึกเก็บรวบรวมโดยเครื่องบันทึกภาพ (Sensors) ที่ติดตั้งบนดาวเทียม เครื่องบันทึกภาพดังกล่าวนี้สามารถเทียบได้กับ Receptor ที่อยู่ในแบบจำลอง ข้อมูลขาเข้า (input) ซึ่งอาจเป็นไปได้ทั้งระบบ MSS หรือ TM อย่างใดอย่างหนึ่ง ข้อมูลขาเข้าทั้งหมดก็คือชุดของค่าสังเกต (observation) หรือจำนวนจุดภาพ (pixel) ทั้งหมด n จุดภาพ ในค่าสังเกตหรือจุดภาพหนึ่งๆ จะประกอบด้วยชุดของค่าระดับสีเทา (Greytone level) ของวัตถุชนิดหนึ่งที่เกิดจากการบันทึกด้วยขนาดช่วงคลื่นหรือแบนด์ (Band or Channel) ต่างๆ เป็น p แบนด์ สำหรับ Categorizer นั้นเทียบได้กับเครื่องมือที่ประกอบด้วยวิธีการทางสถิติต่างๆ ที่ใช้ในการตัดสินใจว่าข้อมูลที่ต้องการจะศึกษาหรือจุดภาพนั้นๆ ว่า ควรเป็นกลุ่มหรือประเภทข้อมูลใด ซึ่งเป็นข้อมูลขาออก (output)



รูปที่ 10 รูปแบบจำลองวิธีการจำแนกข้อมูลจากดาวเทียม

รูปที่ 11 ภาพถ่ายดาวเทียม (Photographic image) ขาว-ดำ ที่ใช้สำหรับแปลภาพด้วยสายตา

3.2 ทฤษฎีสถิติที่เกี่ยวข้อง

วิธีการทางสถิติที่ใช้ในการศึกษาครั้งนี้เกี่ยวข้องและอ้างอิงถึงทฤษฎีสถิติต่างๆ เช่น ทฤษฎีการแจกแจงของตัวแปรสุ่ม (Distribution Theory) ทฤษฎีการตัดสินใจ (Decision Theory) ดังนั้น จำเป็นต้องกล่าวถึงทฤษฎีสถิติดังกล่าวนี้ เพื่อใช้ประกอบในการอธิบายวิธีการทางสถิติต่างๆ ซึ่งจะกล่าวต่อไป

3.2.1 ทฤษฎีการแจกแจงของตัวแปรสุ่ม (Distribution Theory)

ทฤษฎีการแจกแจงของตัวแปรสุ่มที่จะกล่าวถึงนี้ เป็นการเสนอฟังชั่นของตัวแปรสุ่มที่มีการแจกแจงแบบปกติ (Normal Distribution) ที่สำคัญๆ เพื่อใช้อ้างอิงในการนำไปประยุกต์ใช้ต่อไปนับตั้งแต่กรณีของ Univariate Normal Distribution เป็นอันดับแรก

1. Univariate Normal Distribution (1-dimensional)

เมื่อ x มีการแจกแจงแบบปกติ โดยมี pdf เป็น

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\tfrac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] , \quad -\infty < x < \infty \quad \ldots.(1)$$

ซึ่งมีค่าเฉลี่ย (mean)

$$E\left[X\right] = \int_{-\infty}^{\infty} xp(x)dx$$

$$= \mu$$

และความแปรปรวน (variance)

$$E\left[X-\mu\right]^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

$$= \sigma^2$$

2. Bivariate Normal Distribution (2-dimensional)

ให้ $\underset{\sim}{X}$ เป็นเวกเตอร์สุ่ม (random vector) ของ 2 ตัวแปร

$$\underset{\sim}{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

ที่มีการแจกแจง เป็น Bivariate normal distribution มี pdf เป็น

$$p(X) = \frac{1}{2\pi\,\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x_1-\mu_1)}{\sigma_1}^2 - 2\rho\frac{(x_1-\mu_1)}{\sigma_1}\frac{(x_2-\mu_2)}{\sigma_2} + \frac{(x_2-\mu_2)}{\sigma_2}^2\right\}\right],$$

$$-\infty < x_1 , x_2 < \infty \qquad \ldots\ldots\ldots\ldots\ldots(2)$$

เมื่อนิยาม σ_{ii} คือ ความแปรปรวน และ σ_{ij} คือความแปรปรวนร่วม

โดยที่ $i , j = 1,2$

ซึ่งมี $E\left[X_i\right] = \mu_i$

$$\sigma_i^2 = E\left[x_i - \mu_i\right]^2$$

$$= \sigma_{ii}$$

และ ρ คือ สัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ของ x_1 และ x_2

$; -1 < \rho < 1$

$$\rho = \frac{\sigma_{ij}}{\sqrt{\sigma_{11}\sigma_{22}}} \quad ; i \neq j$$

สามารถเขียนได้ในรูปเมตริกซ์ (matrix) ได้ดังนี้

mean vector $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$

covariance vector

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

$$|\Sigma| = \sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}$$

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21}} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{21} & \sigma_{11} \end{bmatrix}$$

$$\sigma_{12} = \rho\sqrt{\sigma_{11}\sigma_{22}} = \sigma_{21}$$

ดังนั้น $p(X) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\mu)' \quad \Sigma^{-1}(X-\mu)\right]$

$$= \frac{1}{2\pi(\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21})^{\frac{1}{2}}} \exp\left[Q \right] \qquad \ldots\ldots\ldots\ldots(3)$$

เมื่อ $Q = \frac{1}{2(\sigma_{11}\sigma_{22} - \sigma_{12}\sigma_{21})} \{\sigma_{22}(x_1 - \mu_1)^2 - (\sigma_{12} + \sigma_{21})(x_1 - \mu_1)(x_2 - \mu_2)$

$+ \sigma_{11}(x_2 - \mu_2)^2\}$

$$= \frac{1}{2(1 - \frac{\sigma_{12}\sigma_{21}}{\sigma_{11}\sigma_{22}})} \{\frac{(x_1 - \mu_1)^2}{\sigma_{11}} - (\frac{\sigma_{12} + \sigma_{21}}{\sigma_{11}\sigma_{22}})(x_1 - \mu_1)(x_2 - \mu_2)$$

$$+ \frac{(x_2 - \mu_2)^2}{\sigma_{22}}\}$$

$$= \frac{1}{2(1 - \frac{\sigma_{12}\sigma_{21})}{\sigma_{11}\sigma_{22}}} \{\frac{(x_1 - \mu_1)^2}{\sigma_{11}} - \frac{(\sigma_{12} + \sigma_{21})}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} \frac{(x_1 - \mu_1)}{\sqrt{\sigma_{11}}} \frac{(x_2 - \mu_2)}{\sqrt{\sigma_{22}}}$$

$$+ (\frac{x_2 - \mu_2)^2}{\sigma_{22}}\}$$

$$= -\frac{1}{2(1 - \rho^2)} \{\frac{(x_1 - \mu_1)^2}{\sigma_1} - 2\rho\frac{(x_1 - \mu_1)}{\sigma_1} \frac{(x_2 - \mu_2)}{\sigma_2}$$

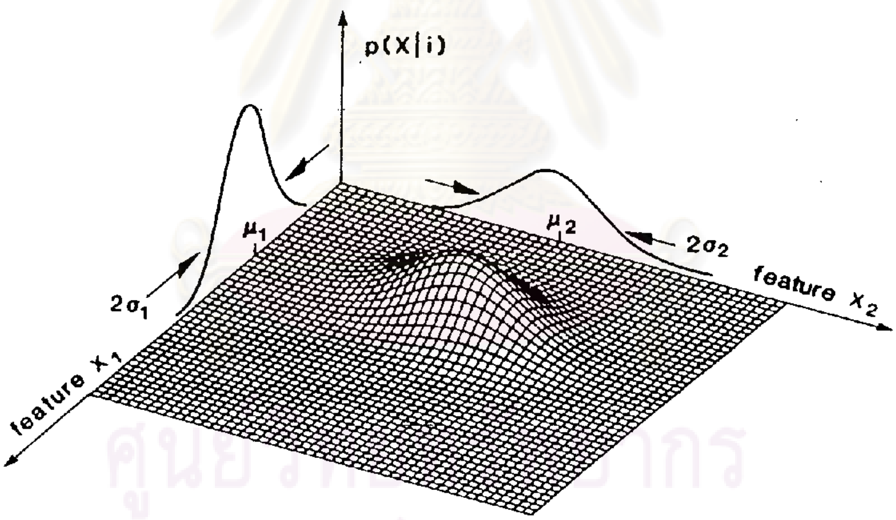
$$+ \frac{(x_2 - \mu_2)^2}{\sigma_2}\}$$

จาก (3) จะได้

$$p(X) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}\left(1 - \frac{\sigma_{12}\sigma_{21}}{\sigma_{11}\sigma_{22}}\right)^{\frac{1}{2}}} \exp\left[Q\right]$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[- \frac{1}{2(1-\rho^2)} \{(\frac{x_1-\mu_1}{\sigma_1}) - 2\rho(\frac{x_1-\mu_1}{\sigma_1})(\frac{x_2-\mu_2}{\sigma_2}) + (\frac{x_2-\mu_2}{\sigma_2})^2\}\right]$$

ซึ่งก็คือ (2) นั่นเอง



mean vector $M = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$

covariance matrix $\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$

correlation coefficient $\rho = \sigma_{12} / (\sigma_1 \sigma_2)^{1/2}$

รูปที่ 12 ลักษณะของ Bivariate Normal Distribution

3. <u>Multivariate Normal Distribution</u> (n-dimensional)

ให้ X เป็นเวคเตอร์สุ่มของ p ตัวแปร

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

ที่มีการแจกแจงเป็น Multivariate normal distribution มี pdf เป็น

$$p(\underset{\sim}{X}) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\tfrac{1}{2}(\underset{\sim}{X}-\mu)'\Sigma^{-1}(\underset{\sim}{X}-u)\right] \quad \ldots\ldots\ldots(4)$$

เมื่อ

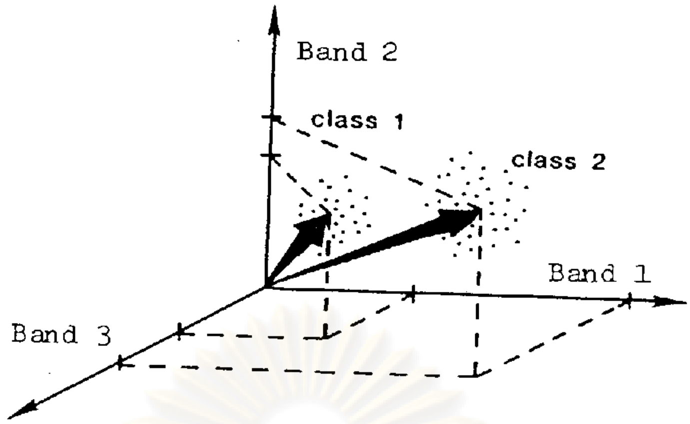
$\underset{\sim}{X}$ คือ p-component column vector

μ คือ p-component mean vector

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = E\left[\underset{\sim}{X}\right]$$

Σ คือ p x p covariance matrix ซึ่งเป็น symmetric positive definite matrix

$$\Sigma = E\left[(\underset{\sim}{X}-u)(\underset{\sim}{X}-\mu)'\right]$$



รูปที่ 13 เวคเตอร์ของจุดภาพในรูป 3 มิติ

3.2.2 ทฤษฎีการตัดสินใจ (Decision Theory)

โดยปกติแล้วการดำเนินกิจกรรมใดๆ ก็ตาม ซึ่งอาจจะเป็นงานที่มีความยุ่งยากสลับซับซ้อน มีหลายขั้นตอน เช่น โครงการอวกาศ หรือแม้แต่การดำเนินชีวิตประจำวันของมนุษย์ ผู้ที่ต้องรับผิดชอบในกิจกรรมนั้นๆ มักจะต้องมีปัญหาการตัดสินใจว่าจะดำเนินการหรือปฏิบัติอย่างไรจึงจะถูกต้อง เหมาะสม นั่นคือ ต้องมีการหาเกณฑ์หรือข้อมูลข่าวสาร มาเป็นตัวกำหนดแนวทางซึ่งก็มักมีแนวทางที่จะต้องเลือกอยู่หลายแนวทาง จึงเกิดปัญหาการตัดสินใจ ที่จะเลือกแนวทางที่ดีที่สุด เกิดการสูญเสียน้อยที่สุดขึ้น ในทฤษฎีการตัดสินใจจะมีปัญหาการตัดสินใจที่ประกอบด้วย กลุ่มสภาวะการณ์ (State of nature) กลุ่มทางเลือก (Action space) และฟังชั่นความสูญเสีย (Loss function) ในการนี้ผู้ตัดสินใจก็จะเลือกทางเลือก (action) ที่ดีที่สุดจากกลุ่มหรือเซต (Set) ทางเลือกทั้งหมดที่มี ซึ่งถ้าทำการเลือกโดยไม่ใช้ข้อมูลอื่นๆมาประกอบก็จะเรียกวิธีการเช่นนี้ว่า Decision without Information และถ้ามีการใช้ข้อมูลอื่นมาประกอบ ซึ่งอาจเป็นการเก็บรวบรวมข้อมูลแบบปฐมภูมิ (Primary data) หรือข้อมูลทุติยภูมิ (Secondary data) ก็ดี การใช้ข้อมูลสนเทศเหล่านี้มาประกอบการตัดสินใจในลักษณะเช่นนี้ เรียกว่า Decision with Information

สำหรับในการศึกษาครั้งนี้มีการเก็บรวบรวมข้อมูล คือ จุดภาพ (pixel) นั่นก็คือ เป็นการใช้ลักษณะการแจกแจงของตัวแปรสุ่มที่มีการแจกแจงความน่าจะเป็นแสดงได้ด้วยฟังชั่นความหนาแน่นที่เรียกว่า Prior probability density function และนำมาประกอบกับหลักเกณฑ์ของทฤษฎีการตัดสินใจเป็นฟังชั่น การตัดสินใจที่ใช้สำหรับการจำแนกประเภทข้อมูลหรือตัดสินใจเลือกจุดภาพที่เก็บรวบรวมมานั้นว่าเป็นวัตถุหรือสสารใดนั่นเอง เรียกฟังชั่นการตัดสินใจลักษณะดังกล่าวนี้ว่า Discriminant function ซึ่งอธิบายได้ดังนี้ (7 : 152)

กำหนดให้

Ω คือ State of nature หรือ Parameter space

$\Omega = (\theta_1, \theta_2, \ldots\ldots \theta_m)$

A คือ Action space

$A = (a_1, a_2, \ldots\ldots a_m)$

$L(a, \theta)$ คือเซต (Set) ของฟังชั่นสูญเสีย (Loss function) ดังนั้น

$L(a_i, \theta_j)$ คือ การสูญเสียที่เกิดขึ้นเมื่อตัดสินใจเลือกเป็น class ที่ i ทั้งๆ ที่เป็น class ที่ j

; $i, j = 1,2,\ldots,m$

ถ้ากำหนดให้ค่าเฉลี่ยของการสูญเสียที่เป็นไปได้ทั้งหมด (Average loss) คือ ความเสี่ยง (Risk) โดยที่

$R_X(a)$ เป็นฟังชั่นการเสี่ยงที่เกิดขึ้น

$$R_X(a_i) = E\left[L(a_i, \theta_j)\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

ด้วยวิธีการตัดสินใจแบบเบย์(Bayes Optimal Decision) จะต้องทำการ Minimize $R_X(a)$ ดังนั้นจาก (5)

$$R_X(a_i) = \sum_{j=1}^{m} L(a_i, \theta_j)\, p(\theta_j | X) \quad \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

โดยที่ $p(\theta_j|X)$ คือ Posterior probability density function of θ_j given X จาก Bayes' rule

$$p(\theta_j|X) = p(X|\theta_j)\, p(\theta_j)/p(X) \quad \ldots\ldots\ldots\ldots\ldots\ldots(7)$$

เมื่อ $p(X|\theta_j)$ คือ Conditional probability density function of X given θ_j

$p(\theta_j)$ คือ Prior probability density function of θ_j

$p(X)$ Marginal probability density function of X ซึ่ง

$$p(X) = \sum_{j=1}^{m} p(X|\theta_j)p(\theta_j)$$

ดังนั้นจาก (6)

$$R_X(a_i) = \sum_{j=1}^{m} L(a_i\ , \theta_j)p(X|\theta_j)p(\theta_j)/p(X) \ldots.(8)$$

จากข้อกำหนดต่อไปนี้จะนำมาประกอบการสร้างฟังชั่นการตัดสินใจต่อไป คือ

(1) การ Mimimize เซตของฟังชั่นใดๆ ก็เปรียบเสมือนการ Maximize ค่าลบของฟังชั่นชุดเดียวกัน ....(9-1)

(2) กำหนดให้เซตของฟังชั่นสูญเสีย (Loss function) เป็น

$$L(a_i\ , \theta_j) = \begin{cases} 0 & \text{เมื่อตัดสินใจถูกต้อง } (i = j) \\ 1 & \text{เมื่อตัดสินใจผิด } (i \neq j) \end{cases} \quad \ldots\ldots\ldots(9\text{-}2)$$

(3) ถ้า $\{g_i(X)\ , i = 1,2,\ldots,m\}$ คือ เซตของฟังชั่นการตัดสินใจใดๆ เราสามารถที่จะทำการ บวก ลบ คูณ ด้วยตัวคงที่ หรือใส่ ln กับฟังชั่นนั้นๆ (Monotonic-transformation) ก็จะไม่ทำให้การตัดสินใจเปลี่ยนแปลง กล่าวคือ

$$g_i^1(X) = g_i(X) \pm k \quad \text{เมื่อ k คือตัวคงที่} \quad \ldots\ldots\ldots\ldots(9\text{-}3)$$

$$g_i^2(X) = g_i(X)\ .\ k \quad \text{เมื่อ k คือตัวคงที่} \quad \ldots\ldots\ldots\ldots(9\text{-}4)$$

$$g_i^3(X) = \ln\ g_i(X) \quad \ldots\ldots\ldots\ldots(9\text{-}5)$$

จาก (8) และ (9-1) ซึ่งเป็นการเปลี่ยนจากการ Minimize เป็น Maximize จะได้ฟังชั่นการตัดสินใจ คือ

$$g_i(X) = - R_x(a_i)$$

$$= - \sum_{j=1}^{m} L(a_i , \theta_j)p(X|\theta_j)p(\theta_j)/p(X)$$

$$= - \sum_{\substack{j=1 \\ i\neq j}}^{m} p(X|\theta_j)p(\theta_j)/p(X)$$

$$= \frac{-1}{p(X)} \sum_{\substack{j=1 \\ j\neq j}}^{m} p(X|\theta_j)p(\theta_j) \qquad \ldots\ldots\ldots\ldots\ldots(10)$$

ในที่นี้ p(X) จะเป็นตัวคงที่ ซึ่งจาก (9-4) จะได้ฟังชั่นการตัดสินใจตัวใหม่ กล่าวคือ

$$g_i'(X) = - \sum_{\substack{j=1 \\ i\neq j}}^{m} p(X|\theta_j)p(\theta_j) \qquad \ldots\ldots\ldots\ldots\ldots(11\text{-}1)$$

จาก $$p(X) = \sum_{j=1}^{m} p(X|\theta_j)p(\theta_j)$$

$$= p(X|\theta_i)p(\theta_i) + \sum_{\substack{j=1 \\ i\neq j}}^{m} p(X|\theta_j)p(\theta_j)$$

$$\sum_{\substack{j=1 \\ i\neq i}}^{m} p(X|\theta_j)p(\theta_j) = p(X) - p(X|\theta_i)p(\theta_i) \qquad \ldots\ldots\ldots\ldots\ldots(11\text{-}2)$$

แทนค่า (11-2) ลงใน (11-1)

$$g_i'(X) = - p(X) + p(X|\theta_i)p(\theta_i)$$

จาก (๑-๒) และ $p(X)$ เป็นตัวคงที่ และจะได้ฟังชั่นการตัดสินใจตัวใหม่ กล่าวคือ

$$g_i''(X) = g_i'(X) + p(X)$$
$$= p(X|\theta_i)p(\theta_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots(12)$$

นั้นคือ ฟังชั่นการตัดสินใจที่ได้จะเท่ากับผลคูณระหว่าง Conditional probability function of X given θ $[p(X|\theta_i)]$ กับ Prior probability density function $[p(\theta_i)]$ และเรียกฟังชั่นการตัดสินใจนี้ว่า Discriminant function โดยจะตัดสินใจเลือก X อยู่ในกลุ่มที่ i ($X \in \theta_i$) ถ้า

$$p(X|\theta_i)p(\theta_i) > p(X|\theta_j)p(\theta_j) \quad ; \quad j = 1,2,\ldots,m$$

นั้นคือ ถ้าให้

$$g_i(X) = p(X|\theta_i)p(\theta_i) \quad \ldots\ldots\ldots\ldots\ldots(13\text{-}1)$$

$$g_j(X) = p(X|\theta_j)p(\theta_j) \quad \ldots\ldots\ldots\ldots\ldots(13\text{-}2)$$

ดังนั้นจะตัดสินใจเลือก X อยู่ใน class ที่ i ก็ต่อเมื่อ

$$g_i(X) > g_j(X) \quad i \neq j \quad \ldots\ldots\ldots\ldots\ldots(13\text{-}3)$$

๓.๓ การประยุกต์

จากทฤษฎีสถิติที่กล่าวแล้วในข้อที่ ๓.๒ เมื่อนำมาประยุกต์ใช้เพื่อการจำแนกประเภทข้อมูลจากดาวเทียมตามวัตถุประสงค์ของการศึกษา อธิบายได้ดังนี้

กำหนดให้

$\Omega = (\theta_1, \theta_2, \ldots, \theta_c)$ เป็นเซต (set) ของกลุ่มหรือประเภทข้อมูล (class) ซึ่งในการศึกษาครั้งนี้แบ่งประเภทข้อมูลออกเป็น ๔ ประเภท (class) คือ ป่าไม้ ปาล์มน้ำมัน นาข้าว และถนน ดังนั้น

Ω = (ป่าไม้, ปาล์มน้ำมัน, นาข้าว, ถนน)

$A = (a_1, a_2, a_3, a_4)$ เป็นเซตของการเลือกหรือการตัดสินใจที่เหมาะสม โดยที่ a_1 คือ การตัดสินใจเลือกจุดภาพ (pixel) ใดๆ ในอาณาบริเวณที่ศึกษาว่าเป็นข้อมูลประเภทป่าไม้ในทำนองเดียวกัน a_4 คือ การตัดสินใจเลือกจุดภาพ (pixel) ในอาณาบริเวณที่ศึกษาว่าเป็นข้อมูลประเภทถนน เป็นต้น

$L(a_i, \theta_j)$ = ฟังชั่นความสูญเสีย (Loss function) ที่เกิดขึ้นเมื่อทำการตัดสินใจเลือก a_i เมื่อกำหนด θ_j ทั้งนี้ $i , j = 1,2,3,4$

$$L(a_i, \theta_j) = \begin{cases} 0 & \text{เมื่อตัดสินใจเลือกจุดภาพนั้นๆ} \\ & \text{เป็นประเภทข้อมูล (class) ที่ถูกต้อง} \\ 1 & \text{เมื่อตัดสินใจเลือกจุดภาพนั้นๆ} \\ & \text{เป็นประเภทข้อมูล (class) ที่ผิด} \end{cases} \quad ......(14)$$

เช่น เกิดความสูญเสีย เมื่อตัดสินใจจุดภาพนั้นให้เป็นปาล์มน้ำมัน โดยความเป็นจริงแล้วเป็นป่าไม้หรือในทำนองกลับกันตัดสินใจให้จุดภาพนั้นเป็นป่าไม้ ทั้งๆ ที่ความเป็นจริงเป็นปาล์มน้ำมัน

X = จุดภาพ (pixel) ที่เป็น 4-component vector ซึ่งแต่ละ component คือ ค่าระดับสีเทาที่อ่านได้จากแต่ละแบนด์ในที่นี้กำหนดใช้จำนวนแบนด์ทั้งหมด 4 แบนด์ ดังนี้

$$(p = 4) \qquad X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

ทั้งนี้ภายใต้สมมติฐาน (assumption) ให้ X มีการแจกแจงเป็นแบบ Multivariate Normal Distribution

$P(X)$ = Marginal pdf ของ X

$h(\theta_j)$ = Prior probability ของ θ_j (หรือความน่าจะเป็นที่จะเกิดประเภทข้อมูลที่ j)

$P(X|\theta_j)$ = Conditional probability density function of X given θ_j

$P(\theta_j|X)$ = Posterior probability density function of θ_j given X

ในทำนองเดียวกันกับ (13-1) และ (13-2) จะได้

$$g_i(X) = P(X|\theta_i)h(\theta_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots(15\text{-}1)$$

$$g_j(X) = P(X|\theta_j)h(\theta_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots(15\text{-}2)$$

ซึ่งเป็นฟังชั่นการตัดสินใจที่ใช้ในการจำแนกประเภทข้อมูลโดยตัดสินใจเลือก Pixel(X) อยู่ในประเภทข้อมูลหรือ class ที่ i ถ้า

$g_i(X) > g_j(X)$ สำหรับทุกค่าของ $i \neq j$ ....(15-3)

โดยสรุปจะเห็นได้ว่า การตัดสินใจของเราจะขึ้นอยู่กับ X ซึ่งเป็นตัวอย่างสุ่ม และ $h(\theta)$ ซึ่งเป็น Prior probability density ของ class ที่เราพิจารณาอยู่เรียกฟังชั่นการตัดสินใจ ($g_i(X)$) นี้ว่า Discriminant function

เมื่อนำทฤษฎีการแจกแจงของตัวอย่างสุ่มที่ได้จาก (1) หรือ (2) หรือ (4) และพิจารณา pdf ของ X เป็นแบบมีเงื่อนไข เมื่อกำหนด θ นั้นก็คือ จะได้ $P(X|\theta_i)$ เป็นไปในแต่ละกรณี กล่าวคือ

ในกรณีของ Univariate Normal Distribution

$$P(X|\theta_i) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{X-\mu_i}{\sigma_i}\right)^2\right] \ldots\ldots\ldots(16)$$

ดังนั้น

$$g_i(X) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{X-\mu_i}{\sigma_i}\right)^2\right] h(\theta_i) \quad \ldots\ldots(17)$$

หรือ ในกรณีของ Multivariate Normal Distribution

$$P(X|\theta_i) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\mu_i)'\Sigma_i^{-1}(X-\mu_i)\right] \ldots\ldots(18)$$

$$g_i(X) = \frac{1}{(2\pi)^{\frac{p}{2}}|\Sigma_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\mu_i)'\Sigma_i^{-1}(X-\mu_i)\right] h(\theta_i) \quad \ldots(19)$$

3.4 การประมาณค่าสถิติจากตัวอย่าง

จาก Discriminant function ที่กล่าวแล้วนั้น (18) จำเป็นต้องใช้ค่าสถิติเกี่ยวกับค่าเฉลี่ย (μ_i) และค่าความแปรปรวนร่วม (Σ_i) ซึ่งไม่สามารถทราบค่าที่แท้จริงของมัน จึงต้องประมาณค่าเหล่านี้จากพื้นที่ตัวอย่าง (Training area) โดยการกำหนดพื้นที่ตัวอย่างที่เหมาะสมในแต่ละประเภทข้อมูล (class) แล้วคำนวณค่าสถิติที่ต้องการจากพื้นที่ตัวอย่างนั้น

กำหนดให้

$$\text{Pixel } X = \begin{bmatrix} x_{i11} \\ x_{i21} \\ x_{i31} \\ x_{i41} \end{bmatrix}$$

x_{ijl} = ค่าระดับสีเทาในพื้นที่ตัวอย่างของจุดภาพที่ l ของแบนด์ที่ j class ที่ i

; $i = 1, 2, 3, 4$

$j = 1, 2, 3, 4$

$l = 1, 2, \ldots, t_i$

เมื่อ t_i = จำนวนจุดภาพทั้งหมดในพื้นที่ตัวอย่างทุกผืนของ class ที่ i

$$m_{ij} = \frac{1}{t_i} \sum_{l=1}^{t_i} x_{ijl} \qquad \ldots\ldots\ldots\ldots\ldots\ldots(20\text{-}1)$$

$$\text{Sample mean vector } M_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{i3} \\ m_{i4} \end{bmatrix}$$

s_{ijk} = ค่าความแปรปรวนร่วมที่คำนวณจากพื้นที่ตัวอย่างระหว่าง Band ที่ j กับ k ของ class ที่ i

$$s_{ijk} = \frac{1}{t_i - 1} \sum_{l=1}^{t_i} (x_{ijl} - m_{ij})(x_{ikl} - m_{ik}) \quad \ldots\ldots\ldots(20\text{-}2)$$

$$\text{Sample covariance matrix } S_i = \begin{bmatrix} s_{i11} & s_{i12} & s_{i13} & s_{i14} \\ s_{i21} & s_{i22} & s_{i23} & s_{i24} \\ s_{i31} & s_{i32} & s_{i33} & s_{i34} \\ s_{i41} & s_{i42} & s_{i43} & s_{i44} \end{bmatrix}$$

โดยวิธีการประมาณค่าแบบ Maximum-Likelihood Estimation จะได้ M_i และ S_i เป็นตัวประมาณที่ไม่เอนเอียง (unbiased estimators) สำหรับ μ_i และ Σ_i ตามลำดับ

สำหรับ Priori probability ($h(\theta_i)$) ซึ่งเป็นความน่าจะเป็นของการเกิดประเภทข้อมูล (class) ใดๆ ตามที่กำหนดนั้น สามารถคำนวณได้ดังนี้

1. Equal priori probability คือให้ความน่าจะเป็นในการที่จะเกิด class ใดๆ เท่ากัน ดังนั้น

ถ้าให้ C = จำนวนประเภทข้อมูลทั้งหมดที่ได้กำหนด

$$h(\theta_1) = h(\theta_2) = \ldots = h(\theta_i) = \frac{1}{C}$$

2. Priori probability proportional to size คือให้ความน่าจะเป็นในการที่จะเกิด class ใดๆ ไม่เท่ากัน ทั้งนี้โดยการพิจารณาข้อมูลที่ได้จากแหล่งอื่นประกอบ เช่น จากผลการสำรวจที่ผ่านมา (Previous survey) ตลอดจนจากประสบการณ์ที่พอจะทราบขนาดของเนื้อที่ของประเภทข้อมูลใดๆ ในท้องที่หรือพื้นที่ที่ต้องการจะศึกษา

ดังนั้น Discriminant function ที่จะนำมาใช้ในการจำแนกข้อมูลจาก ดาวเทียม คือ จาก (19) จะเป็น

$$g_i(X) = \frac{1}{(2\pi)^{\frac{p}{2}}|S_i|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-M_i)'S_i^{-1}(X-M_i)\right] h(\theta_i) \quad \ldots(21)$$

3.5 <u>วิธีการทางสถิติที่ใช้จำแนกประเภทข้อมูล</u>

จากวิธีการจำแนกประเภทข้อมูลจากดาวเทียมทั้งหมดที่กล่าวแล้วในข้อที่ 3.1 นั้น กำหนดใช้วิธีการจำแนกประเภทข้อมูลจากดาวเทียมด้วยเครื่องคอมพิวเตอร์เป็นแบบ Supervised analysis สำหรับการศึกษาครั้งนี้ จากความสามารถของโปรแกรมสำเร็จรูป MOA-RECOGX มีวิธีการทางสถิติที่ใช้จำแนกประเภทข้อมูลจากดาวเทียม 3 วิธีด้วยกัน นับตั้งแต่การใช้วิธีการและสมมติฐานทางสถิติขั้นพื้นฐานจนถึงการสถิติชั้นสูง ทั้งนี้ขึ้นอยู่กับข้อสมมติฐานที่กำหนดใช้ในแต่ละวิธี ได้แก่ Level Slicing, Euclidean distance และ Maximum-likelihood ทั้งนี้ในแต่วิธีการมีหลักเกณฑ์ ดังต่อไปนี้

3.5.1 <u>Level Slicing (LS)</u> เป็นวิธีการทางสถิติขั้นพื้นฐาน ใช้หลักเกณฑ์ทางสถิติง่ายๆ ประกอบกับการพิจารณาตัดสินใจของผู้วิจัย (subjective) ที่จะเลือกใช้ข้อมูลข่าวสาร (information) ใดๆ อาทิเช่น จากประสบการณ์ของผู้วิจัยเองหรือผลการสำรวจทางภาคพื้นดิน (Ground Truth) ตลอดจนผลการวิจัยแหล่งอื่นๆ ที่ผ่านมา เป็นต้น ตัวอย่างหลักเกณฑ์ทางสถิตินั้นอาจพิจารณาได้จาก

(1) กราฟแผนภูมิแท่ง (Histogram) ที่แสดงความถี่ของระดับสีเทาในพื้นที่ตัวอย่าง (Training area) ศึกษาลักษณะการกระจายของค่าระดับสีเทาของประเภทข้อมูล (class) แต่ละประเภทในแต่ละแบนด์แล้วพิจารณาเลือกค่าระดับสีเทาสูงสุด (maximum) และต่ำสุด (minimum) จากกราฟแผนภูมิแท่ง ทั้งนี้พิจารณาว่าค่าระดับสีเทาของแต่ละประเภทข้อมูลมีความเหลื่อมซ้อน (overlap) กันมากน้อยเพียงใด ถ้าเหลื่อมซ้อนกันจะทำให้ค่าสูงสุดและค่าสูดของแต่ละประเภทข้อมูลคู่หนึ่งๆ เป็นค่าเดียวกันหรือใกล้เคียงกัน ทำให้พิจารณาตัดสินใจเลือกจุดภาพกระทำไม่ได้จำเป็นต้องปรับค่าสูงสุดหรือต่ำสุดเสียใหม่ ตัวอย่างเช่น จากรูปที่ 14 อาจพิจารณากำหนดค่าสูงสุดเท่ากับ 45 ค่าต่ำสุดเท่ากับ 20 เป็นต้น

รูปที่ 14 แผนภูมิแท่ง ใช้พิจารณาค่าสูงสุดและต่ำสุดในวิธีการแบบ Level Slicing

(2) ค่าสถิติต่างๆ ที่ได้จากพื้นที่ตัวอย่าง (Training area) ของกลุ่มหรือประเภทข้อมูล (class) ที่กำหนดไว้ ค่าสถิตินั้นได้แก่ ค่าเฉลี่ยของระดับสีเทา และค่าเบี่ยงเบนมาตรฐาน เมื่อนำค่าสถิติดังกล่าวนี้มากำหนดค่าระดับสีเทาสูงสุดและต่ำสุด จากสมการ

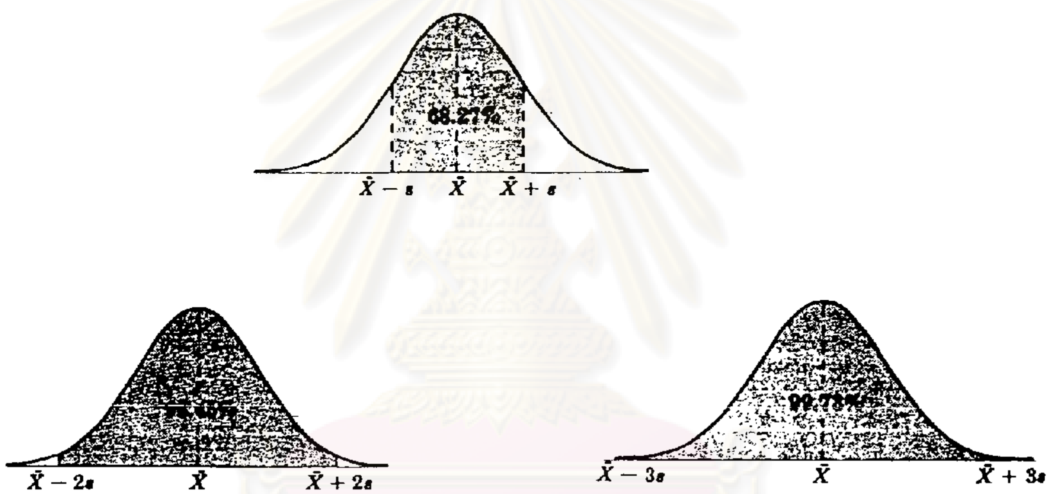
ค่าระดับสีเทาสูงสุด = ค่าเฉลี่ยของระดับสีเทา + (k) ค่าเบี่ยงเบนมาตรฐาน ....(22)

ค่าระดับสีเทาต่ำสุด = ค่าเฉลี่ยของระดับสีเทา - (k) ค่าเบี่ยงเบนมาตรฐาน ....(23)

โดยอาศัยทฤษฎีการแจกแจงตัวอย่าง (Sampling distribution) ที่กล่าวว่าตัวอย่างสุ่มจากประชากรที่มีการแจกแจงเป็นแบบปกติแล้ว การแจกแจงของค่าเฉลี่ยตัวอย่างของทุกครัวเรือนตัวอย่างที่เป็นไปได้ทั้งหมด ก็จะมีการแจกแจงเป็นแบบปกติด้วย ดังนั้นค่าเฉลี่ยตัวอย่างที่เป็นไปได้จะเกิดภายใต้พื้นที่เส้นโค้งปกติด้วยความน่าจะเป็นต่างๆ กัน

เมื่อพิจารณาเฉพาะจุดภาพ ที่อยู่ภายใต้พื้นที่เส้นโค้งปกติด้วยความน่าจะเป็นต่างๆ กัน กล่าวคือ กำหนดให้เท่ากับ 0.68 (หรือ k = 1) , 0.95 (หรือ k = 2) และ 0.99 (หรือ k = 3) ผลจากกำหนดค่าความน่าจะเป็นดังกล่าวมากน้อยแตกต่างกัน เปรียบได้เสมือนกับการกำหนดระดับความมั่นใจ (coefficient of confidence) ที่ใช้ในการประมาณค่าแบบช่วง คือ ถ้ากำหนดระดับความมั่นใจต่ำก็จะทำให้ช่วงการประมาณค่าแคบ แต่ให้ค่าประมาณที่ใกล้

ค่าจริง (parameter) มากขึ้น ในทำนองเดียวกัน เมื่อกำหนดระดับความมั่นใจสูงก็จะทำให้ช่วงการประมาณค่ากว้าง แต่ให้ค่าประมาณที่ห่างไปจากค่าจริง เมื่อนำเปรียบ-เทียบกับจุดภาพของแต่ละประเภทข้อมูลที่อยู่ใต้พื้นที่เส้นโค้งปกติ ซึ่งอาจจะเหลื่อมซ้อนกัน (overlap) ทำให้ส่วนที่เหลื่อมซ้อนกันนั้น ไม่สามารถตัดสินใจได้ว่า จุดภาพนั้นๆ เป็นข้อมูลประเภทใด จึงไม่นำจุดภาพที่เหลื่อมซ้อนกันนั้นมาพิจารณา โดยตัดออกจากการตัด-สินใจ เลยทำให้จุดภาพที่ได้จะเสนอผลออกมาในลักษณะที่เรียกว่าไม่สามารถจำแนกได้ (unclassification)



รูปที่ 15 พื้นที่ใต้เส้นโค้งปกติ ส่วนที่แรเงาหมายถึงพื้นที่ใต้เส้นโค้งปกติ เมื่อ k = 1, 2 และ 3

ค่าระดับสีเทาสูงสุดและต่ำสุดของประเภทข้อมูลใดๆ ในแต่ละแบนด์จะเป็นช่วงที่ใช้เป็นเกณฑ์พิจารณาจำแนกข้อมูลทั้งหมดของบริเวณที่ศึกษา ข้อมูลหรือจุดภาพใดที่อยู่ในช่วงค่าสูงสุดและต่ำสุดครบทุกแบนด์ ก็จะพิจารณาตัดสินใจให้จุดภาพนั้นเป็นประเภทข้อมูลนั้น ตัวอย่าง เช่น เมื่อนำค่าระดับสีเทาของจุดภาพในพื้นที่ตัวอย่างของประเภทข้อมูลปาล์มน้ำมันมาคำนวณหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานได้ในแต่ละแบนด์ ดังนี้

	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน
แบนด์ที่ 4	48.40	2.99
แบนด์ที่ 5	52.78	4.64
แบนด์ที่ 6	81.30	4.91
แบนด์ที่ 7	99.71	7.02

สมมติว่าจากการศึกษากราฟแผนภูมิแท่งประกอบข้อมูลภาคพื้นดิน พบว่าค่าระดับสีเทาของจุดภาพประเภทข้อมูลปาล์มน้ำมัน เหลื่อมซ้อนกับข้อมูลประเภทป่าไม้มาก จึงพิจารณาจุดภาพเฉพาะพื้นที่ใต้โค้งปกติ เป็นร้อยละ 68.27 ของพื้นที่ใต้เส้นโค้งปกติของจุดภาพทั้งหมด นั้นคือ กำหนดใช้ $k = 1$ คำนวณหาค่าสูงสุดและต่ำสุดของประเภทข้อมูลปาล์มน้ำมันในแต่ละแบนด์ได้เป็น (51.39 , 45.41) (57.42 , 48.14) (86.21 , 76.39) (106.73 , 92.69) ดังนั้น เกณฑ์การตัดสินใจ เลือกจุดภาพใดๆ ว่า เป็นข้อมูลประเภทใด เช่น จุดภาพ x_1 มีค่าระดับสีเทาเป็น 47.25, 50.29, 78.92 และ 100.45 ในแบนด์ที่ 4, 5, 6, และ 7 ตามลำดับ จะเห็นว่า จุดภาพ x_1 มีค่าระดับสีเทาอยู่ในช่วงระหว่างค่าสูงสุดและต่ำสุดในทุกแบนด์ ดังนั้นจะตัดสินใจ เลือกจุดภาพ x_1 นั้นเป็น ปาล์มน้ำมัน

3.5.2 Maximum-likelihood (ML) เป็นวิธีการทางสถิติที่นำมาใช้ในการจำแนกประเภท เพื่อคัดเลือกตัวแปรกลุ่มหนึ่งหรือชุดหนึ่งที่ต้องการศึกษาออกเป็นประเภทข้อมูลใดๆ ตามที่กำหนดไว้ โดยการนำ Discriminant function (21) แทนค่าในฟังชั่นดังกล่าวด้วยค่าสถิติต่างๆ เช่น ค่าเฉลี่ย และเมตริกความแปรปรวนร่วม ตลอดจน Prior probability ของการที่เกิดประเภทข้อมูลใดๆ คำนวณได้ค่าจำนวนหนึ่งใช้เป็นเกณฑ์ในการตัดสินใจ กล่าวคือ จากทฤษฎีการตัดสินใจของเบยส์ (Optimal Bayes Decision) จะพิจารณาเลือกจุดภาพหนึ่งๆ อยู่ในประเภทข้อมูลที่ i ถ้า

$$g_i(X) > g_j(X) \quad \text{สำหรับทุกค่าของ } i \neq j$$

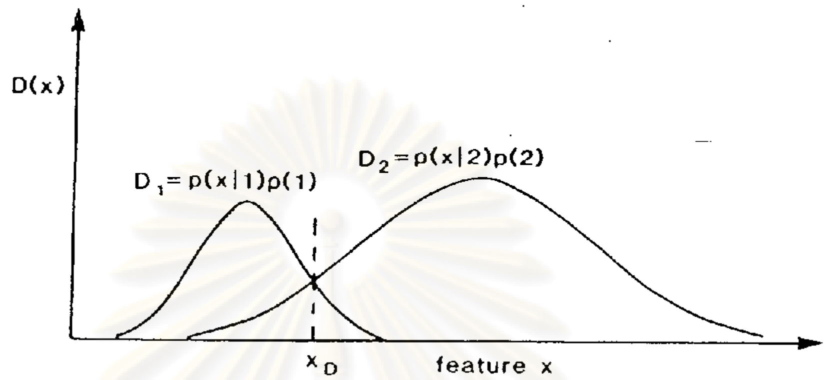
นั้นคือถ้า กำหนดให้มีประเภทข้อมูลอยู่ 2 ประเภท (2 class)

จุดภาพนั้นอยู่ในประเภทข้อมูลที่ 1 ถ้า $g_1(X) > g_2(X)$

หรือ จุดภาพนั้นอยู่ในประเภทข้อมูลที่ 2 ถ้า $g_2(X) > g_1(X)$

โดยที่ $g_i(X) = P(X/\theta_i)h(\theta_i)$

อาจแสดงด้วยภาพ 1 มิติ ในกรณีสมมติว่า มี 2 ประเภทข้อมูล ดังนี้



รูปที่ 16 **แสดง** Discriminant function **ของ** Optimal Bayes Decision

ด้วยคุณสมบัติของทฤษฎีการตัดสินใจ เราสามารถที่จะ บวก ลบ คูณ หรือใส่ ln ด้วยตัวคงที่กับ Discriminant function ใดๆ แล้วจะไม่มีผลต่อการตัดสินใจ ตัวอย่างเช่น

$$g_i(X) = a\left[P(X|\theta_i)h(\theta_i)\right] + b \qquad \ldots\ldots\ldots\ldots(24\text{-}1)$$

$$g_i(X) = \ln\left[P(X|\theta_i)h(\theta_i)\right] \qquad \ldots\ldots\ldots\ldots(24\text{-}2)$$

ซึ่งฟังชั่นทั้งสอง ก็ยังคงเป็น Discriminant function ที่ใช้ในการตัดสินใจอยู่นั่นเอง จาก (21)

$$g_i(X) = \ln|h(\theta_i)| - \frac{p}{2}\ln(2\pi) - \tfrac{1}{2}\ln|S_i| - \left[\tfrac{1}{2}(X-M_i)'S_i^{-1}(X-M_i)\right] \qquad \ldots\ldots\ldots\ldots(25)$$

จะเห็นว่า $-\frac{p}{2}\ln(2\pi)$ จะเป็นตัวคงที่ซึ่งปรากฏอยู่ในทุกๆ $g_i(X)$ ทำให้สามารถตัดออกจากสมการได้ โดยไม่ทำให้ผลการตัดสินใจไม่เปลี่ยนแปลง ดังนั้น ผลที่เหลือก็คือ

$$g_i(X) = \ln|h(\theta_i)| - \tfrac{1}{2}\ln|S_i| - \tfrac{1}{2}(X-M_i)' S_i^{-1}(X-M_i) \quad \ldots(26)$$

เรียกสมการที่ได้นี้ว่า Maximum-likelihood ซึ่งเป็น Discriminant function ที่ใช้ในการพิจารณาตัดสินใจเลือกจุดภาพ (pixel) ใดๆ

3.5.3 Euclidean distance (ED) เป็นการนำ Discriminant function มาใช้ในการจำแนกประเภทเช่นเดียวกันกับในวิธี Maximum-likelihood โดยพิจารณาเฉพาะระยะทางของจุดภาพต่างๆ ที่กระจายอยู่ในรูปหลายมิติ จุดภาพใดที่มีระยะทางใกล้กับค่าเฉลี่ยของประเภทข้อมูลนั้นๆ มากที่สุดก็จะตัดสินใจเลือกจุดภาพใดเป็นประเภทข้อมูลนั้น

เมื่อพิจารณากรณีพิเศษ (special case) ของวิธี Maximum-likelihood กล่าวคือ (9 : 49)

1. ถ้าเราสมมติเมตริกของภาพความแปรปรวนร่วม (covariance matrix) ของทุกๆ ประเภทข้อมูล (class) เท่ากัน

$$\Sigma_i = \Sigma_j = \ldots = \Sigma_o \quad \ldots\ldots\ldots\ldots\ldots(27)$$

นั่นคือ

$$S_i = S_j = \ldots = S_o \quad \ldots\ldots\ldots\ldots\ldots(28)$$

2. กำหนดให้ priori probability ของการเกิดประเภทข้อมูลนั้นๆ เท่ากันทุกๆ class

$$h(\theta_i) = h(\theta_j) = \ldots = h(\theta_o) \quad \ldots\ldots\ldots\ldots\ldots(29)$$

ดังนั้น Discriminant function จาก (26) จะเปลี่ยนเป็น

$$g_i(X) = A - \frac{1}{2}(X-M_i)' S_o^{-1}(X-M_i) \quad \ldots\ldots\ldots\ldots\ldots(30)$$

ซึ่ง A จะเป็นตัวคงที่และเท่ากับ

$$A = \ln\left|h(\theta_o)\right| - \frac{1}{2}\ln\left|S_o\right| \quad \ldots\ldots\ldots\ldots\ldots(31)$$

สามารถตัด A ออกจากสมการได้ ดังนั้น จาก (30) จะเปลี่ยนเป็น

$$g_i'(X) = -\frac{1}{2}(X-M_i)' S_o^{-1}(X-M_i) \quad \ldots\ldots\ldots\ldots\ldots(32)$$

เรียกฟังชั่นที่ได้นี้ว่า Mahalanobis distance

เมื่อพิจารณาให้ค่าระดับสีเทาของจุดภาพ (pixel) ในแต่ละแบนด์ไม่มีความสัมพันธ์กัน (uncorrelated) ดังนั้น ความแปรปรวนร่วมจะมีค่าเป็นศูนย์ ทำให้

$$\Sigma_i = \begin{bmatrix} \sigma_{i11} & & & \\ & \sigma_{i22} & & \\ & & \sigma_{i33} & \\ & & & \sigma_{i44} \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots(33)$$

ดังนั้น

$$S_i = \begin{bmatrix} S_{i11} & & & \\ & S_{i22} & & \\ & & S_{i33} & \\ & & & Si44 \end{bmatrix} \quad \ldots\ldots\ldots\ldots\ldots\ldots(34)$$

จาก (32) จะได้

$$g_i^1(X) = -\frac{1}{2}(X-M_i)' S_i^{-1}(X-M_i)$$

หรือจาก (9-1) และ (9-2) จะได้

$$g_i^2(X) = (X-M_i)' S_i^{-1}(X-M_i)$$

$$g_i^2(X) = -\sum_{j,k=1}^{4} \frac{(x_j - m_{ij})^2}{s_{ijj}} \quad ; \quad i \neq j \ \ldots\ldots(35)$$

จะเรียก (35) นี้ว่า Euclidean distance ซึ่งเป็น Discriminant function ที่ใช้พิจารณาเลือกจุดภาพ (pixel) X อยู่ใน class ที่ i เมื่อ

$$g_i(X) > g_j(X) \quad ; \text{ สำหรับทุกค่าของ } i \neq j$$

กล่าวโดยสรุป จะเห็นว่า Discriminant function จาก (26) และ (35) จะใช้ข้อมูลข่าวสาร (information) ที่ได้ ดังนี้

Maximum-likelihood จะใช้ข้อมูลข่าวสารเกี่ยวกับ

1. ค่าเฉลี่ย (mean vector) : μ_i
2. ค่าความแปรปรวนร่วม (covariance matrix) : Σ_i
3. Priori probability : $h(\theta_i)$

จะพิจารณาเลือก pixel X อยู่ใน class ที่ i ถ้า

$$g_i(X) > g_j(X) \qquad \text{สำหรับทุกค่าของ } i \neq j$$

ซึ่งเป็น Discriminant function ที่ใช้ ความน่าจะเป็นของการเกิด class หนึ่งๆ ของกลุ่มหรือชุดของข้อมูลทั้งหมดที่ศึกษา เป็นส่วนประกอบของฟังชั่นดังกล่าวนี้ ซึ่งอาจกล่าวได้ว่า ถ้า Priori probability ของ class ใดสูง ก็จะทำให้ $g_i(X)$ มีค่าสูงตามไปด้วย

Euclidean distance จะใช้ข้อมูลข่าวสารเกี่ยวกับ

1. ค่าเฉลี่ย (mean vector) : μ_i
2. ค่าระดับสีเทาของจุดภาพในแต่ละแบนด์ ไม่มีความสัมพันธ์ (uncorrelated) ดังนั้น ความแปรปรวนร่วม = 0 จึงใช้แต่ค่าความแปรปรวน (Variance)

จะพิจารณาเลือก Pixel X อยู่ใน class ที่ i ถ้า

$$g_i(X) > g_j(X) \qquad \text{สำหรับทุกค่าของ } i \neq j$$

ซึ่งเป็น Discriminant function ที่พิจารณาถึงระยะทางระหว่างค่าสังเกตกับค่าเฉลี่ย