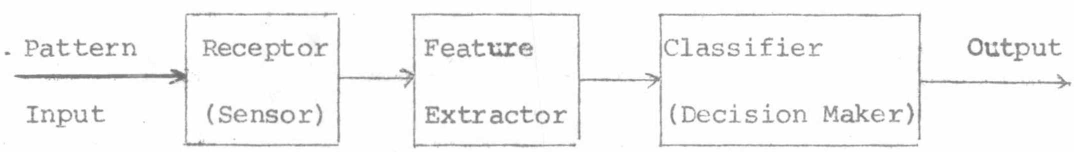
บทที่ 3

ทฤษฎีการจำแนกภาพ

ที่มาของทฤษฎี

ขั้นตอนการวิเคราะห์ข้อมูลจากดาวเทียมที่สำคัญ ก็คือการ Classification หรือการจำแนกประเภทข้อมูลที่มีอยู่แต่ละจุดว่าเป็นอะไร เช่น เป็นป่า สวน หรือน้ำ เป็นต้น โดยมีเครื่องมือที่นิยมนำมาใช้ช่วยในการตัดสินใจเพื่อเลือกประเภทที่ต้องการก็คือ วิชา Pattern Recognition ซึ่งมีรูปแบบการจำแนกข้อมูลคือ (รูปที่ 9)



รูปที่ 9 ระบบการจำแนกข้อมูล

ระบบการจำแนกข้อมูลตามรูปที่ 9 นี้ เมื่อนำมาใช้กับข้อมูลที่ได้รับจากดาวเทียมแล้ว อาจกล่าวได้ว่า Receptor หรือ Sensor ก็คือกล้องถ่ายภาพระบบ RBV หรือระบบ MSS ที่ถ่ายภาพพื้นผิวโลก ข้อมูลที่ได้จากการถ่ายภาพเรียกว่า Ground Resolution Element (GRE) GRE แต่ละจุดจะมี d จำนวน ตามจำนวนกล้องในระบบถ่ายภาพนั้น ๆ เช่น ถ้าเป็นระบบ 4 กล้อง GRE แต่ละจุดจะมี 4 แบนด์ จากนั้น Feature Extractor ซึ่งอาจเป็นเทคนิคหรือวิธีการที่เหมาะสมจะทำหน้าที่แปลงค่า GRE แต่ละจุดให้เป็นค่าใน class ต่าง ๆ เพื่อให้ Classifier หรือ Decision Maker ทำหน้าที่ตัดสินว่า GRE นั้น ๆ ควรจะอยู่ใน class อะไร

วิธีการที่จะช่วยในการตัดสินใจนั้นมีหลายวิธี เช่น Bayes Decision และ Maximum Likelihood Decision เป็นต้น ในผลงานของ H.N. Mahabala, et. al.[1] ซึ่งเป็นชาวอินเดีย ก็ใช้ Maximum Likelihood ในงานวิจัยเพื่อจำแนกข้อมูลจากดาวเทียม ซึ่งอาจกล่าวได้ว่าวิธีการนี้เหมาะแก่การทำแผนที่การใช้ที่ดิน (Land-use map) โดยปกติข้อมูลจะมีจำนวนมากคือประมาณ 31 ล้านจุดต่อภาพ (ทั้ง 4 แบนด์) ทำให้การคำนวณแม้จะใช้เครื่องคอมพิวเตอร์ก็ยังต้องใช้เวลามาก ดังนั้นการทำ sampling pixels คือการเลือกตัวอย่างข้อมูลบางจุดเพื่อนำมาจำแนกข้อมูล จะช่วยลดจำนวนข้อมูลลงได้

DISCRIMINANT FUNCTIONS

ในการคำนวณค่าซึ่งจะนำมาใช้เป็นตัวแทนหรือตัวกลางเพื่อนำมาใช้ตัดสินใจนั้นมีหลายอย่าง แต่สิ่งหนึ่งที่จะนำมาใช้ในวิทยานิพนธ์นี้ก็คือ Discriminant Functions ต่อไปนี้จะอธิบายหลักการใช้ discriminant functions

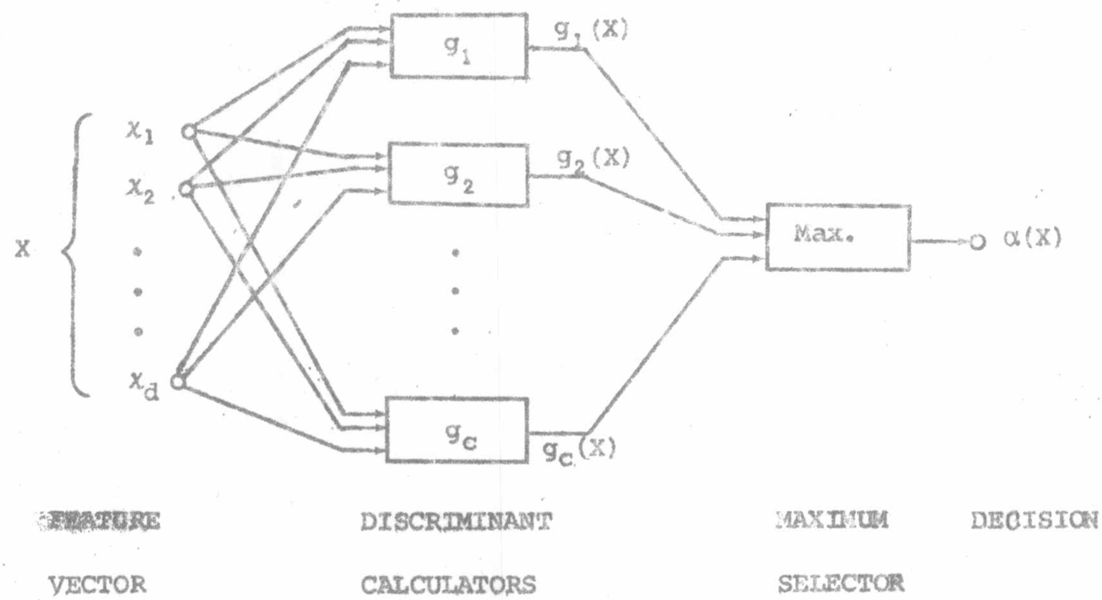
สมมุติให้มี class ที่ต้องการจำแนกข้อมูลอยู่ c class ให้ $g_i(X)$ เป็น discriminant function ของ class ที่ i , $i = 1,\ldots,c$
และ feature X จะอยู่ใน class ω_i ถ้า

$$g_i(X) > g_j(X) \qquad \text{สำหรับทุก } j \neq i \tag{1}$$

นั่นคือเราจะคำนวณค่า discriminant ทั้งหมด c ค่า แล้วจึงเลือก class ซึ่งให้ค่า discriminant สูงสุด (maximize $g_i(X)$) เป็น class ที่ต้องการ ดังรูปที่ 10

---

[1] H. N. Mahabala, et al. "Bulk processing of ERTS Imagery," (1973): 669-682

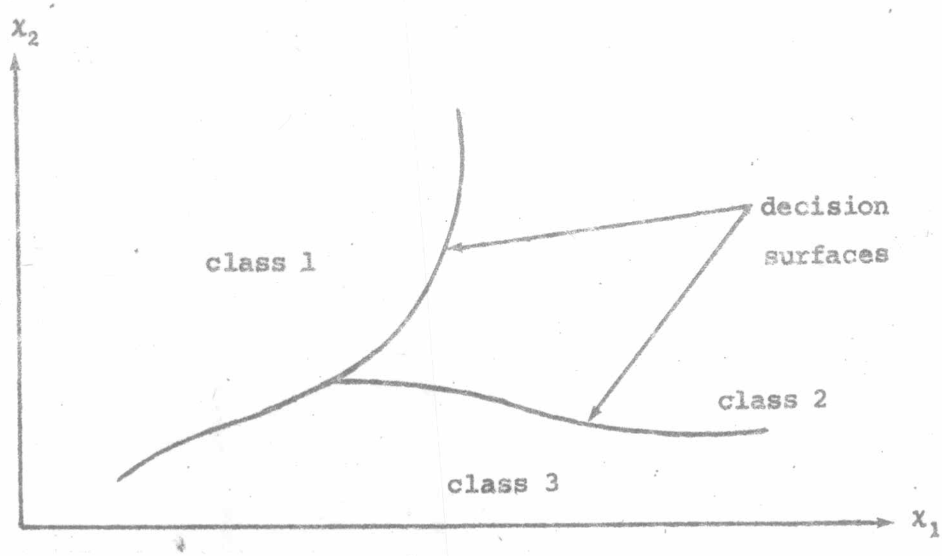


รูปที่ 10 Pattern Classifer

ในทางคณิตศาสตร์ GRE แต่ละจุดสามารถแทนด้วย vector ซึ่งเรียกว่า feature vector ใน d-dimensional space และ feature space จะถูกแบ่งออกเป็นส่วน ๆ ตามจำนวน class ที่เรามี ซึ่งเรียกว่า decision regions แต่ละพื้นที่จะถูกแบ่งออกจากกันด้วย decision surfaces หรือ decision boundaries (ดังรูปที่ 11) decision surface นี้ สามารถแทนได้ด้วยสมการ

$$g_i(X) = g_j(X) \tag{2}$$

ที่จุด surface นี้ เราไม่สามารถตัดสินใจให้แน่นอนลงไปได้ว่า feature X จะอยู่ใน class ใด แต่ก็มักตัดสินให้ X อยู่ใน ω_i ถ้า $g_i(X) = g_j(X)$



รูปที่ 11 Decision Regions และ Decision Surfaces

ในการสร้าง discriminant functions นั้น เราอาศัยหลักพิจารณาจาก loss ที่เกิดขึ้น โดยกำหนดให้

$\Omega = \{\omega_1,\ldots,\omega_c\}$ เป็น เซท (set) ของ class ต่าง ๆ หรือที่เรียกว่า state of nature

$A = \{\alpha_1,\ldots,\alpha_c\}$ เช่น เซท ของการตัดสินใจ หรือที่เรียกว่า action

$\lambda(\alpha_i|\omega_j)$ เป็น loss ที่เกิดขึ้นสำหรับการเลือก action α_i เมื่อ state of nature ที่แท้จริงคือ ω_j หรือก็คือ loss ที่เกิดขึ้นเมื่อเลือก class α_i ทั้ง ๆ ที่ class ที่แท้จริงคือ ω_j

และ X เป็น feature vector ที่มี d-component

การตัดสินใจของเราก็คือ เลือก action ที่ให้ค่า expected loss ที่ต่ำที่สุด ซึ่งจะเป็น Optimum Bayes decision

ดังนั้นสำหรับ feature vector X ที่กำหนดให้ expected loss function (หรือ risk และเป็น conditional risk) จะเขียนในรูปของสมการคณิตศาสตร์ได้ดังนี้

$$R(\alpha_i|X) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|X) \qquad (3)$$

โดยที่ $P(\omega_j|X)$ คือ a posteriori probability และจะคำนวณได้โดยอาศัยกฎของ Bayes (Bayes Rule) ซึ่งกล่าวว่า

$$P(\omega_j|X) = \frac{p(X|\omega_j)P(\omega_j)}{p(X)} \qquad (4)$$

เมื่อ

$$p(X) = \sum_{j=1}^{c} p(X|\omega_j)P(\omega_j) \qquad (5)$$

โดยที่[1] $p(X|\omega_j)$ คือ probability density function ของ X given ω_j

$P(\omega_j)$ คือ a priori probability ของ ω_j

การกำหนดค่า loss เพื่อหาค่า conditional risk นั้น ยากแก่การปฏิบัติมาก ดังนั้นจึงสมมุติให้เป็น equal loss และการพิจารณาค่า equal loss ที่น่าสนใจเป็นอย่างยิ่งก็คือ กรณีที่เรียกว่า Zero-one loss function หรือ symmetrical กล่าวคือ

ให้

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 & i=j \\ 1 & i\neq j \end{cases} \qquad i,j=1,\ldots,c \qquad (6)$$

คือกำหนดให้ loss เป็น 0 เมื่อเลือก class ได้ถูกต้อง

และ loss เป็น 1 เมื่อเลือก class ผิด

---

[1] ในที่นี้จะใช้สัญลักษณ์ p เมื่อกล่าวถึง probability density function และ P เมื่อกล่าวถึง probability mass function

ดังนั้น

$$R(\alpha_i|X) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|X)$$

$$= \sum_{\substack{j=1 \\ j\neq i}}^{c} P(\omega_j|X)$$

$$= 1-P(\omega_i|X) \qquad (7)$$

ตามกฎการตัดสินใจของ Bayes (Bayes decision rule) คือเลือก action ที่ให้ค่า risk ต่ำสุด ซึ่งเท่ากับเป็นการ minimize $R(\alpha_i|X)$ หรือก็คือ maximize-$R(\alpha_i|X)$ หรืออาจกล่าวได้ว่าเลือก action ที่ให้ค่า a posteriori probability สูงสุดนั่นเอง และกฎการตัดสินใจ (dicision rule) ก็คือเลือก ω_i ถ้า $P(\omega_i|X) > P(\omega_j|X)$ สำหรับทุก $j \neq i$

ดังนั้นจึงได้ discriminant ที่เหมาะสมคือ

$$g_i(X) = -R(\alpha_i|X) \qquad (8)$$

เมื่อค่า discriminant function ที่ใหญ่ที่สุดก็คือค่า risk ที่น้อยที่สุดนั่นเอง

ดังนั้น $$g_i(X) = P(\omega_i|X)-1 \qquad (9)$$

สำหรับเซท (set) ของ discriminant functions ใด ๆ นั้นเราอาจหาเซทของ discriminant functions อื่น ซึ่งเป็น monotonically increasing function มาแทนได้ โดยที่ผลการจำแนกข้อมูลไม่เปลี่ยนแปลง

เช่นให้ $g_i(X)$; $i = 1,\ldots,c$ เป็นเซทของ discriminant functions ใด ๆ ดังนั้น

$$g_i'(X) = g_i(X) + \text{constant} \qquad i=1,2,\ldots,c$$

$$g_i''(X) = \log[g_i(X)] \qquad i=1,2,\ldots,c$$

ก็จะเป็นเซทของ discriminant functions ที่ใช้แทนกันได้ เป็นต้น

ดังนั้นจากสมการ (9) จึงสามารถเขียน discriminant functions ใหม่ได้ดังสมการข้างล่าง และจะเห็นได้ว่าบางสมการยังทำให้การคำนวณและการทำความเข้าใจนั้นง่ายขึ้นอีกด้วย

$$g_i(X) = P(\omega_i|X) \tag{10}$$

จากสมการ (4) และ (10) ได้

$$g_i(X) = \frac{p(X|\omega_i)P(\omega_i)}{p(X)} \tag{11}$$

ค่า $p(X)$ เป็นค่าซึ่งไม่ใช่ฟังก์ชั่นของ i ฉะนั้นสามารถตัดออกจากสมการโดยผลการตัดสินใจจะไม่เปลี่ยนแปลง

$$g_i(X) = p(X|\omega_i)P(\omega_i) \tag{12}$$

$$g_i(X) = \log p(X|\omega_i)+\log P(\omega_i) \tag{13}$$

สำหรับสมการเหล่านี้ กฎการตัดสินใจ ก็ยังเหมือนเดิมกล่าวคือ

เลือก X อยู่ใน ω_i ถ้า $g_i(X) > g_j(X)$ สำหรับทุก $j \neq i$

กฎการเลือกนี้ก็คือ MAXIMUM LIKELIHOOD DECISION RULE นั่นเอง

จะเห็นว่า discriminant functions ที่เราต้องการนั้นขึ้นอยู่กับค่า

$p(X|\omega_i)$ คือ conditional probability density function (p.d.f.)

และ $P(\omega_i)$ คือ a priori probability

ค่าทั้งสองนี้เราไม่สามารถทราบค่าที่แท้จริงของมันได้ แต่เราอาจกำหนดหรือประมาณค่าได้ เช่น กำหนดหรือประมาณค่า a priori probability ของแต่ละ class ขึ้นมาจากประสบการณ์ในอดีตโดยที่ $\Sigma P(\omega_i)$ จะต้องเท่ากับ 1 ส่วนค่า conditional p.d.f. นั้นเราอาจหาค่าของมันได้โดยการสมมุติลักษณะ distribution ของ X ที่เรากำลังสนใจขึ้นมา แล้วจึงประมาณค่า parameters ที่มีความสัมพันธ์กับ distribution นั้น วิธี

นี้เรียกว่า "parametric"

DISCRIMINANT FUNCTIONS กับ NORMAL DENSITY

สมมุติฐานที่สำคัญและเป็นที่นิยมคือสมมุติให้ $p(X|\omega_i)$ มีการกระจายแบบ Multivariate Normal Distribution เนื่องจากข้อมูลในแต่ละภาพและในแต่ละแบนด์ มีจำนวนมากคือประมาณ 8 ล้านจุด ดังนั้นจึงอาศัย Central-Limit Theorem ในการสมมุติลักษณะของการกระจายเป็น normal distribution ได้ ซึ่งอาจกล่าวว่า probability distribution จะเข้าใกล้ normal เมื่อจำนวน sample เข้าใกล้ infinity

และก่อนที่จะได้กล่าวถึงกรณี multivariate เพื่อให้เป็นที่เข้าใจยิ่งขึ้นจะขอกล่าวตั้งแต่กรณี univariate normal density ซึ่งเป็นพื้นฐานเบื้องต้นของ distribution นี้ก่อน

Univariate Normal Density (1-dimensional)

ซึ่งมีสูตรโดยทั่วไปของ normal density ดังนี้

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-u}{\sigma}\right)^2\right] \tag{14}$$

โดยที่
$$u = E[x] \tag{15}$$

และ
$$\sigma^2 = E[(x-u)]^2 \tag{16}$$

เป็น mean และ variance ตามลำดับ และเขียนเป็นสัญญลักษณ์แทนได้ ดังนี้ $p(x) \sim N(u,\sigma^2)$ และกล่าวว่า x มีการกระจายแบบ normal ด้วย mean = u และ variance = σ^2

ในการหาค่า discriminant functions นั้นก็คือ สมมุติให้ $p(x|\omega_i) \sim N(u_i, \sigma_i^2)$ โดยที่ u_i และ σ_i^2 เป็น mean และ variance ของแต่ละ class ที่ i ซึ่งเขียนได้ว่า

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x-u_i}{\sigma_i}\right)^2\right] \tag{17}$$

จากสมการ (13) จึงได้ discriminant functions ดังนี้

$$g_i(x) = -\frac{1}{2}\left(\frac{x-u_i}{\sigma_i}\right)^2 - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma_i^2 + \log p(\omega_i) \tag{18}$$

จะเห็นว่า $-\frac{1}{2}\log 2\pi$ เป็นค่าคงที่ และจะปรากฏอยู่ในทุก ๆ $g_i(x)$ ทำให้สามารถตัดออกจากสมการได้ โดยที่การตัดสินใจไม่เปลี่ยนแปลง ผลที่เหลือก็คือ

$$g_i(x) = -\frac{1}{2}\left(\frac{x-u_i}{\sigma_i}\right)^2 - \frac{1}{2}\log\sigma_i^2 + \log P(\omega_i) \tag{19}$$

<u>Bivariate Normal Density</u> (2-dimensional)

ให้ X เป็น vector ของ 2 ตัวแปร

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

และมีสูตรโดยทั่วไปของ density ดังนี้

$$p(X) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{q}{2}\right] \tag{20}$$

เมื่อ

$$q = \frac{1}{1-\rho^2}\left[\left(\frac{x_1-u_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-u_1}{\sigma_1}\right)\left(\frac{x_2-u_2}{\sigma_2}\right) + \left(\frac{x_2-u_2}{\sigma_2}\right)^2\right] \tag{21}$$

โดยที่

$$u_1 = E[x_1] \qquad u_2 = E[x_2] \tag{22}$$

$$\sigma_{ii} = E\left[(x_i-u_i)\right]^2 \quad ; \quad \sigma_i^2 = \sigma_{ii} \tag{23}$$

และ ρ เป็น correlation coefficient ; $-1 < \rho < 1$

และสามารถเขียนในเทอมของ mean vector และ covariance matrix ได้ดังนี้

$$\mu = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{24}$$

และ $$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \quad \text{หรือ} \quad \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{bmatrix} \qquad (25)$$

เมื่อ $\sigma_{12} = \rho\sigma_1\sigma_2 = \sigma_{21}$ และ $\sigma_1^2 = \sigma_{11}$, $\sigma_2^2 = \sigma_{22}$

และ density functions เขียนใหม่ในเทอมของ matrix ได้ดังนี้

$$p(X) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\mu)^t\Sigma^{-1}(X-\mu)\right] \qquad (26)$$

เมื่อ $|\Sigma|$ คือ determinant ของ Σ

และ $(X - \mu)^t$ คือ transpost ของ $(X - \mu)$

สำหรับรูป matrix นี้ สามารถขยายออกไปจนถึง d-demensional ดังในหัวข้อต่อไป

<u>Multivariate Normal Density</u> (d-dimensional)

สูตรโดยทั่วไปของ multivariate normal density นี้ มีดังนี้คือ

$$p(X) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(X-\mu)^t\Sigma^{-1}(X-\mu)\right] \qquad (27)$$

เมื่อ X คือ d-component column vector

μ คือ d-component mean vector

Σ คือ $d\times d$ covariance matrix

Σ^{-1} คือ inverse ของ Σ

โดยที่

$$\mu = E[X] \qquad (28)$$

และ $$\Sigma = E\left[(X-\mu)(X-\mu)^t\right] \qquad (29)$$

ในการหาค่า discriminant functions นั้นเราจะสมมุติให้ $p(X|\omega_i) \sim N(\mu_i, \Sigma_i)$ และในทำนองเดียวกันจะได้สมการ discriminant functions ดังนี้

$$g_i(X) = -\frac{1}{2}(X-\mu_i)^t\Sigma_i^{-1}(X-\mu_i) - \frac{1}{2}\log|\Sigma_i| + \log P(\omega_i) \qquad (30)$$

ซึ่งจะตรงกับกฎการตัดสินใจ คือ

เลือก X อยู่ใน ω_i ถ้า $g_i(X) > g_j(X)$ สำหรับทุก $j \neq i$

<u>การประยุกต์ในทางปฏิบัติ</u>

โดยปกติค่า u_i, σ_i^2 หรือ μ_i, Σ_i นั้นเราไม่ทราบค่าที่แท้จริงของมัน จะต้องประมาณค่าที่ต้องการเหล่านี้ขึ้นมาจาก training samples โดยสร้าง training areas ที่เหมาะสมกับแต่ละ class ขึ้นมา แล้วคำนวณค่า parameters ที่ต้องการจาก training samples โดยอาศัยวิธีการประมาณค่าแบบ Maximum Likelihood Estimate

สำหรับ univariate : unbiased estimate[1] สำหรับ u_i และ σ_i^2 คือ

$$\hat{u}_i = \frac{1}{n_t}\sum_{j=1}^{n_t} x_{ij} \qquad (31)$$

$$\hat{\sigma}_i^2 = \frac{1}{n_t-1}\sum_{j=1}^{n_t}(x_{ij}-\hat{u}_i)^2 \qquad (32)$$

เมื่อ n_t คือ จำนวน training samples ของ class ที่ i

และกรณี multivariate : unbiased estimate สำหรับ μ_i และ Σ_i คือ

$$\hat{\mu}_i = \frac{1}{n_t}\sum_{j=1}^{n_t} X_{ij} \qquad (33)$$

$$\hat{\Sigma}_i = \frac{1}{n_t-1}\sum_{j=1}^{n_t}(X_{ij}-\hat{\mu}_i)(X_{ij}-\hat{\mu}_i)^t \qquad (34)$$

เมื่อ n_t คือจำนวน training samples ของ class ที่ i

---

[1] ถ้า A เป็นตัวแปรใด ๆ กล่าวว่า $\hat{A}$ จะเป็น unbiased estimate ของ A ถ้า $E(\hat{A}) = A$

ในทางปฏิบัติ feature vector X คือข้อมูลเกี่ยวกับพื้นผิวโลกที่ได้จากการบันทึกภาพด้วยกล้องระบบ RBV หรือ MSS บนดาวเทียมสำรวจทรัพยากร ดังนั้นหากเป็นระบบ MSS feature vector X จะมี 4 components (4 multispectral bands) แต่ละ component ก็คือ video data ที่มีค่าอยู่ระหว่าง 0-63 นั่นเอง

ดังนั้นสามารถเขียน feature vector X ได้ ในรูป

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

ส่วนค่า $\hat{\mu}_i$ และ $\hat{\Sigma}_i$ ซึ่งเป็น unbiased estemate ของ μ_i และ Σ_i นั้นจะประมาณค่าได้จาก training samples ของแต่ละ class โดยใช้สูตรในการคำนวณค่าดังสมการที่ (33) และ (34) ตามลำดับ ซึ่งสามารถเขียนเป็นแต่ละ compenent ได้ดังนี้

$$\hat{\mu}_i = M_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ m_{i3} \\ m_{i4} \end{bmatrix} \qquad \text{เมื่อ } m_{ij} = \frac{1}{n_t} \sum_{\ell=1}^{n^t} x_{ij\ell}$$

และ

$$\hat{\Sigma}_i = S_i = \begin{bmatrix} s_{i11} & s_{i12} & s_{i13} & s_{i14} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ s_{i41} & s_{i42} & s_{i43} & s_{i44} \end{bmatrix}$$

เมื่อ

$$s_{ijk} = \frac{1}{n_t - 1} \sum_{\ell=1}^{n_t} (x_{ij\ell} - m_{ij})(x_{ik\ell} - m_{ik})$$

$i = 1, 2, \ldots, c$; c คือจำนวน class

$j, \; k = 1, 2, 3, 4$

และ n_i คือจำนวน training samples ใน class ที่ i

ส่วนค่า $P(\omega_i)$ ซึ่งเป็น a priori probability นั้นก็คือโอกาสที่จะเป็น class ใด class หนึ่ง ซึ่งโดยปกติเป็นการยากที่จะกำหนดค่า probability ที่แท้จริงนี้ได้ จึงใช้สมมุติฐานว่า ให้ทุก ๆ class มีโอกาสเกิดเท่า ๆ กัน นั่นก็คือสมมุติ ให้เป็น equal a priori probability นั่นเอง

ดังนั้น $P(\omega_1) = P(\omega_2) = \ldots = P(\omega_i) = \frac{1}{c} \quad i = 1, \ldots, c$

ทำให้ $P(\omega_i) = \frac{1}{c}$ ทุก ๆ $g_i(X)$ จึงสามารถตัดออกจากสมการได้

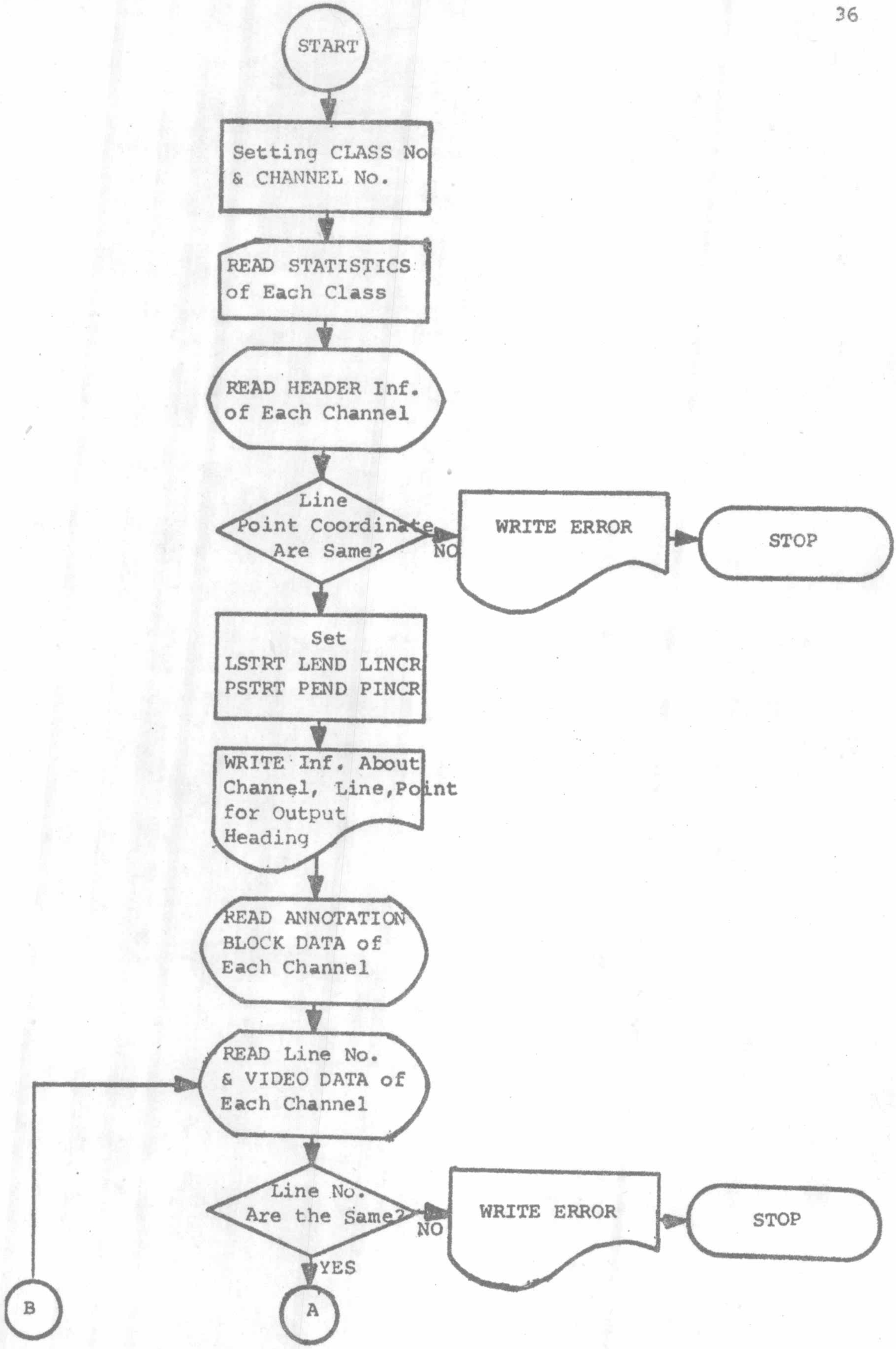
ดังนั้น จากสมการ (30) ค่า discriminant functions ที่ใช้งานจริง ๆ คือ

$$g_i(X|\omega_i) = -\frac{1}{2}(X-M_i)^t S_i^{-1}(X-M_i) - \frac{1}{2}\log|S_i| \qquad (35)$$

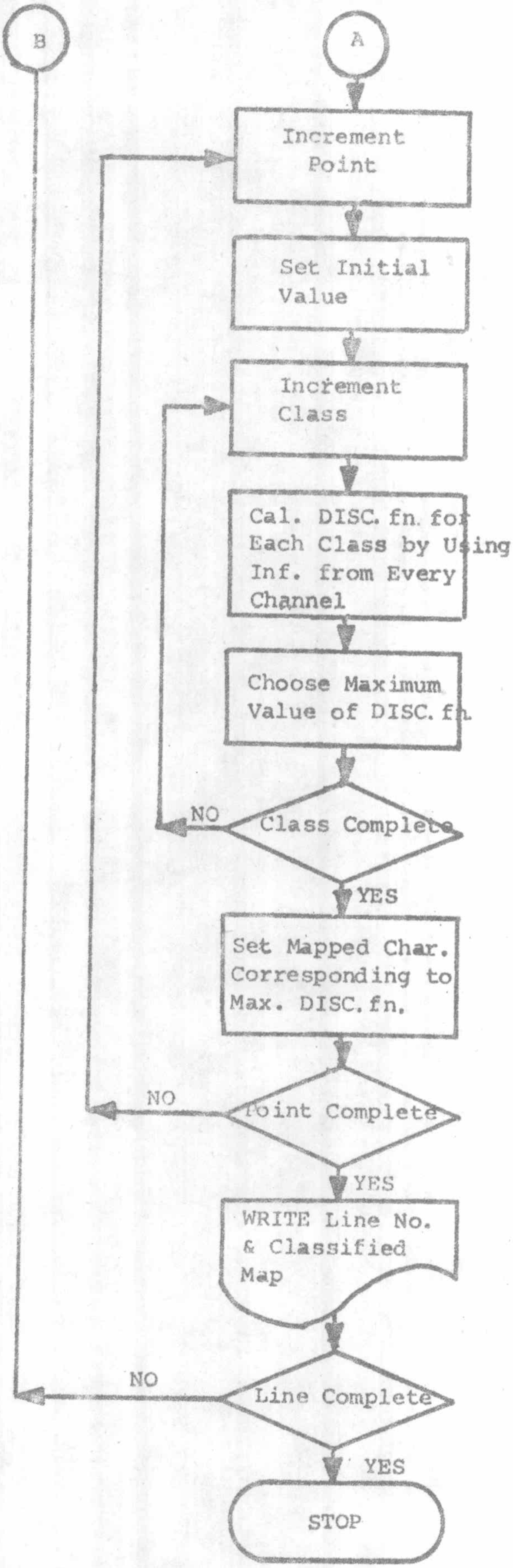
จากสมการ discriminant functions ที่ (36) นี้ เมื่อนำมาเขียนโปรแกรมเพื่อจำแนกข้อมูลแล้ว มีรายละเอียดของ flow-chart ดังรูปที่ 12

ในการคำนวณค่า discriminant functions นี้ จะต้องคำนวณทีละจุดและในแต่ละจุดจะต้องหา $g_i(X)$ ถึง c ค่า แล้วจึงเลือก class ซึ่งมีค่า discriminant สูงสุด เมื่อเป็นดังนี้จึงทำให้การคำนวณเสียเวลามากแม้จะใช้เครื่องคอมพิวเตอร์ก็ตาม ซึ่งเป็นปัญหาใหญ่มากสำหรับวิธีการนี้ การแก้ปัญหาทำได้ดังนี้คือ

1. เนื่องจาก feature X เป็น vector 4 components จาก 4 multi-spectral bands ลักษณะของข้อมูลในแต่ละแบนด์อาจจะแตกต่างกันหรือคล้ายคลึงกัน (ดังรายละเอียดในบทที่ 1) ทำให้การศึกษาลักษณะข้อมูลนั้นไม่จำเป็นต้องใช้หมดทั้งสี่แบนด์ ดังนั้นถ้าเราสามารถเลือกแบนด์ที่เกี่ยวข้องกัน เพื่อนำมาเป็นตัวแทนในการจำแนกข้อมูลแทนที่จะใช้ข้อมูลจากทุกแบนด์ โดยที่แบนด์ที่เราจะเลือกนั้น จะต้องมีความสัมพันธ์กับลักษณะที่เราต้องการจำแนก เช่น ในที่นี้ลักษณะเด่นที่เราต้องการจำแนกข้อมูลจะเป็นป่าไม้ สวน ฯลฯ



รูปที่ 12 ผังภาพแสดงขั้นตอนการจำแนกภาพด้วยวิธี Maximum Likelihood



รูปที่ 12 (ต่อ) ผังภาพแสดงขั้นตอนการจำแนกภาพด้วยวิธี Maximum Likelihood

ซึ่งเป็นพืชพรรณที่มีสีเขียว ข้อมูลจากแบนด์ 5 และ 7 สามารถใช้เป็นตัวแทนของทั้ง 4 แบนด์ได้ เมื่อเป็นอย่างนี้ จะเห็นว่า เวลาในการคำนวณก็จะสามารถลดลงได้

2. การจำแนกข้อมูลถ้าทำทุก ๆ จุด ถึงแม้ว่าจะให้รายละเอียดชัดเจนแต่ก็ใช้เวลามาก การจำแนกเพียงบางจุด โดยพยายามไม่ให้เสียความชัดเจน หรือรายละเอียดไปมากนัก จะช่วยลดเวลาลงได้ เช่น จำแนกข้อมูลทุก ๆ จุดเว้นจุด ทุก ๆ บรรทัดเว้นบรรทัด จะลดเวลาลงได้ถึง 4 เท่า