Statistical Classification

- Main Assumption: The prototypes (and unknown) are drawn from an underlying statistical distribution.

- Bayes Decision Theory: Assumes all statistics are known.

- Parametric: Probability density functions (p.d.f's) are known or assumed. Parameters are unknown.

- Nonparametric: Estimate p.d.f.'s or use other statistical techniques.

Bayes Decision Theory

- Based on quantifying the tradeoffs between various decisions using probability and the costs that accompany such decisions
- All of the relevant probability values are known.

Outline

- Introductory Example
- Bayes Decision Rule for Minimum Error (2-Class)
- Bayes Decision Rule for Minimum Risk (2-Class)
- Discriminant Functions
- Bayes Minimum Error – Multiple Classes
- Bayes Minimum Risk – Multiple Classes
- Special Cases
- Review of Bayes Decision Theory
- Mahalanobis Distance
- Example – Bayes Minimum Error for Normal Density
- Error Probabilites and Likelihood Ratio
- Probability of Error: Examples

Introductory Example

Aerial images

Classify – land, water

S_1=land

S_2=water

-> Feature Extractor -> Classifier ->

A priori probabilities:

$p(S_1)=0.3$

$p(S_2)=0.7$

Prior probabilities reflect our prior knowledge of how likely we are to get S_1 or S_2.

If unknown,

Crude estimates of probabilities:

$$P(S_k) = \frac{M_k}{\sum_{k=1}^{K} M_k}$$

Decision Rule

- Decide S_1 iff $P(S_1)>P(S_2)$
- This rule makes sense if we are to judge just one fish, but not for many fish

Features

- Let's use feature information to improve the classifier
- $x_1 \propto I_{blue}$
- $x_2 \propto$ T="feature regularity"
- $\underline{x}$=feature vector
- $p(x_1|S_2)=p(I_{blue}|\text{water})$=class-conditional p.d.f.
- That is the probability density function for x given that the state of nature is S_1
- $p(\underline{x}|S_i)$=class-conditional p.d.f. (or state-conditional p.d.f) also likelihood of S_1 w.r.t. $\underline{x}$
- Bayes Decision Theory: assume $p(\underline{x}|S_i)$ are known and that $P(S_i)$ are known.

Review: Bayes Statistics and Probability

Bayes Formula

$p(S_1|\underline{x})$=?=a posteriori probability

$$P(S_k \mid \underline{x}) = \frac{p(\underline{x} \mid S_k)P(S_k)}{p(\underline{x})}, \; k=1,2$$

- In English

Posterior = (Likelihood x Prior) / Evidence

Likelihood => $p(\underline{x}|S_k)$

Prior => $p(S_k)$

Evidence => $p(\underline{x}) = \sum_{k=1}^{2} p(\underline{x} \mid S_k)P(S_k)$

- Bayes formula converts the prior probability $p(S_k)$ to a posteriori probability (or posterior) $p(S_k|\underline{x})$: the probability of the state of nature being S_k given that feature value $\underline{x}$ has been measured.
- $p(\underline{x}|S_k)$=the likelihood of S_k with respect to $\underline{x}$, a term to indicate the category S_k for which $p(\underline{x}|S_k)$ is large is more likely to be the true category.
- The posterior depends on the likelihood and the prior as the product.

Bayes Decision Rule for Minimum Error (2-class)

$p(S_1|\underline{x}) > p(S_2|\underline{x}) \Rightarrow \underline{x} \in S_1$

$p(S_2|\underline{x}) > p(S_1|\underline{x}) \Rightarrow \underline{x} \in S_2$

Use Bayes Theorem

$$P(S_k \mid \underline{x}) = \frac{p(\underline{x} \mid S_k)P(S_k)}{p(\underline{x})}$$

$$p(\underline{x}) = \sum_{k=1}^{K} p(\underline{x} \mid S_k)P(S_k)$$

Bayes decision rule for minimum error (2-class)

$p(\underline{x}|S_1)p(S_1) > p(\underline{x}|S_2)p(S_2) \Rightarrow \underline{x} \in S_1$

$p(\underline{x}|S_1)p(S_1) < p(\underline{x}|S_2)p(S_2) \Rightarrow \underline{x} \in S_2$

Rearrange:

$$\frac{p(\underline{x} \mid S_1)}{p(\underline{x} \mid S_2)} > \frac{P(S_2)}{P(S_1)} \Rightarrow \underline{x} \in S_1$$

$$< \quad \Rightarrow \underline{x} \in S_2$$

Likelihood ratio, $l(\underline{x}) = \frac{p(\underline{x} \mid S_1)}{p(\underline{x} \mid S_2)}$

$\frac{P(S_2)}{P(S_1)} = T$ = threshold value

Log Likelihood ratio

$h(\underline{x}) = -\ln[\, l(\underline{x})\,] = \ln[\, p(\underline{x}|S_2)\,] - \ln[\, p(\underline{x}|S_1)\,] < \ln[\, P(S_1)/P(S_2)\,] \Rightarrow \underline{x} \in S_1$

Error probabilities

Def. of Prob. Error => $P_e = p(S_1)\int_{\Gamma 2} p(\underline{x}|S_1)dx + p(S_2)\int_{\Gamma 1} p(\underline{x}|S_2)dx$

Minimize P_e

Integrands ≥ 0 always

⇨ assign $\underline{x}$ to S_1 when $p(\underline{x}|S_2)P(S_2) < p(\underline{x}|S_1)P(S_1)$

Bayes minimum error classifier

Γ_1: $p(S_1|\underline{x}) > P(S_2|\underline{x})$

Γ_2: $p(S_1|\underline{x}) < P(S_2|\underline{x})$

Bayes Decision Rule for Minimum Risk (2-class)

Generalize Bayes minimum error.

- Cost of different classification errors may be different.
- Let $C(S_k|S_j)$=Cost of classifying $\underline{x}$ as S_k when it should be S_j.

$C(S_k|S_j)=C_{kj}$ in DHS λ_{kj} (page 24)
$C(S_2|S_1) > C(S_1|S_1)$ or $C_{21} > C_{11}$
$C(S_1|S_2) > C(S_2|S_2)$ or $C_{12} > C_{22}$

$$\underline{C}=\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

Conditional average loss or risk = $R(S_k|\underline{x})$
$R(S_1|\underline{x})=C(S_1|S_1)P(S_1|\underline{x}) + C(S_1|S_2)P(S_2|\underline{x})$
$R(S_2|\underline{x})=C(S_2|S_1)P(S_1|\underline{x}) + C(S_2|S_2)P(S_2|\underline{x})$

Decision rule – take the action that minimizes the total expected risk.

Total expected risk:

$$R= \int_{\Gamma_1} R(S_1|\underline{x})p(\underline{x})d\underline{x} + \int_{\Gamma_2} R(S_2|\underline{x})p(\underline{x})d\underline{x}$$

Choose Γ_1 and Γ_2 so R is minimum, i.e., so $R(S_k|\underline{x})$ is minimum for each $\underline{x}$.
⇨ decide S_1 if $R(S_1|\underline{x}) < R(S_2|\underline{x})$
⇨ decide S_2 if $R(S_1|\underline{x}) > R(S_2|\underline{x})$

Decision Rule (Minimum Risk Classfier)

$$R(S_1|\underline{x}) \underset{S_2}{\overset{S_1}{\lessgtr}} R(S_2|\underline{x})$$

Express in terms of known densities and C_{ij}:

$$C_{11}P(S_1|\underline{x}) + C_{12}P(S_2|\underline{x}) \overset{<}{\underset{>}{}} C_{21}P(S_1|\underline{x}) + C_{22}P(S_2|\underline{x})$$

$$(C_{11}-C_{21})P(S_1|\underline{x}) \overset{<}{\underset{>}{}} (C_{22}-C_{12})P(S_2|\underline{x})$$

$$P(S_k|\underline{x}) = P(S_k)p(x|S_k)/p(\underline{x})$$

$$(C_{21}-C_{11})p(\underline{x}|S_1)P(S_1) \overset{>}{\underset{<}{}} (C_{12}-C_{22})p(\underline{x}|S_2)P(S_2)$$

Likelihood ratio, $l(\underline{x}) = \dfrac{p(\underline{x}\,|\,S_1)}{p(\underline{x}\,|\,S_2)} \overset{>}{\underset{<}{}} \dfrac{(C_{12}-C_{22})}{(C_{21}-C_{11})}\dfrac{P(S_2)}{P(S_1)}$

Note: if $C_{11}=0$, $C_{12}=1$, $C_{21}=1$, $C_{22}=0$

$l(\underline{x}) = \dfrac{p(\underline{x}\,|\,S_1)}{p(\underline{x}\,|\,S_2)} \overset{>}{\underset{<}{}} \dfrac{P(S_2)}{P(S_1)}$ => Bayes minimum error rule.