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)
- Discriminant Functions
- Bayes Minimum Error – Multiple Classes
- 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

$p(x)$: probability density function (p.d.f) for x

$p(y)$: p.d.f. for y

Joint p.d.f. $p(x,y)=p(x)p(y)$ if random variable x and y are statistically independent.

If x and y are dependent (i.e., y gives a better estimate of x), the conditional probability of x given y,

$p(x|y) = p(x,y)/p(y)$

if x and y are independent, then $p(x|y)=p(x)$, therefore $p(x,y)=p(x)p(y)$.

$p(y|x)=p(x,y)/p(x)$, $p(x,y)=p(y|x)p(x)$

then $p(x|y)=[p(y|x)p(x)]/p(y)$ it is called Bayes formula.

Let's rewrite this,

$p(S|x)=[p(x|S)p(S)]/p(x)$

where

$p(x|S)$ is the likelihood of S with respect to x.

$p(s)$ is the prior density

$p(x)$ is called evident, a scalar scaling factor.

$p(S|x)$ is the posterior probability (i.e., probability of being S given the feature x)

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, $l(\underline{x}) = \dfrac{p(\underline{x} \mid S_1)}{p(\underline{x} \mid S_2)}$

$\dfrac{P(S_2)}{P(S_1)} = T$ = 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: $p(S_1|\underline{x}) < P(S_2|\underline{x})$