ANLY-601
Advanced Pattern Recognition

Spring 2018

L02 – Bayes Classifier

GEORGETOWN UNIVERSITY

Bayesian Decision Theory

- Two classes ω_1, ω_2
- Class priors P_1, P_2

- Class-conditional densities $p(x \mid \omega_i)$ or *likelihood*

- Posteriors by Bayes rule

$$P(\omega_i \mid x) = P_i \, p(x \mid \omega_i) / p(x)$$

where the unconditional density is

$$p(x) = P_1 \, p(x \mid \omega_1) + P_2 \, p(x \mid \omega_2)$$

Posterior

The posterior probability

$$P(\omega_i \mid x) = P_i \; p(x \mid \omega_i) \,/\, p(x)$$

is proportional to the

prior * likelihood

$$P(\omega_1 \mid x) + P(\omega_2 \mid x) = \frac{P_1 \; p(x \mid \omega_1)}{p(x)} + \frac{P_2 \; p(x \mid \omega_2)}{p(x)} = \frac{p(x)}{p(x)} = 1$$

Proportionality constant insures normalization

Bayes Decision Rule

Seems intuitive to choose the most likely class, given the feature measurement vector x and the class priors

$$If\ \ P(\omega_1|x) > P(\omega_2|x),\ \ choose\ \omega_1.\ \ Otherwise\ choose\ \omega_2.$$

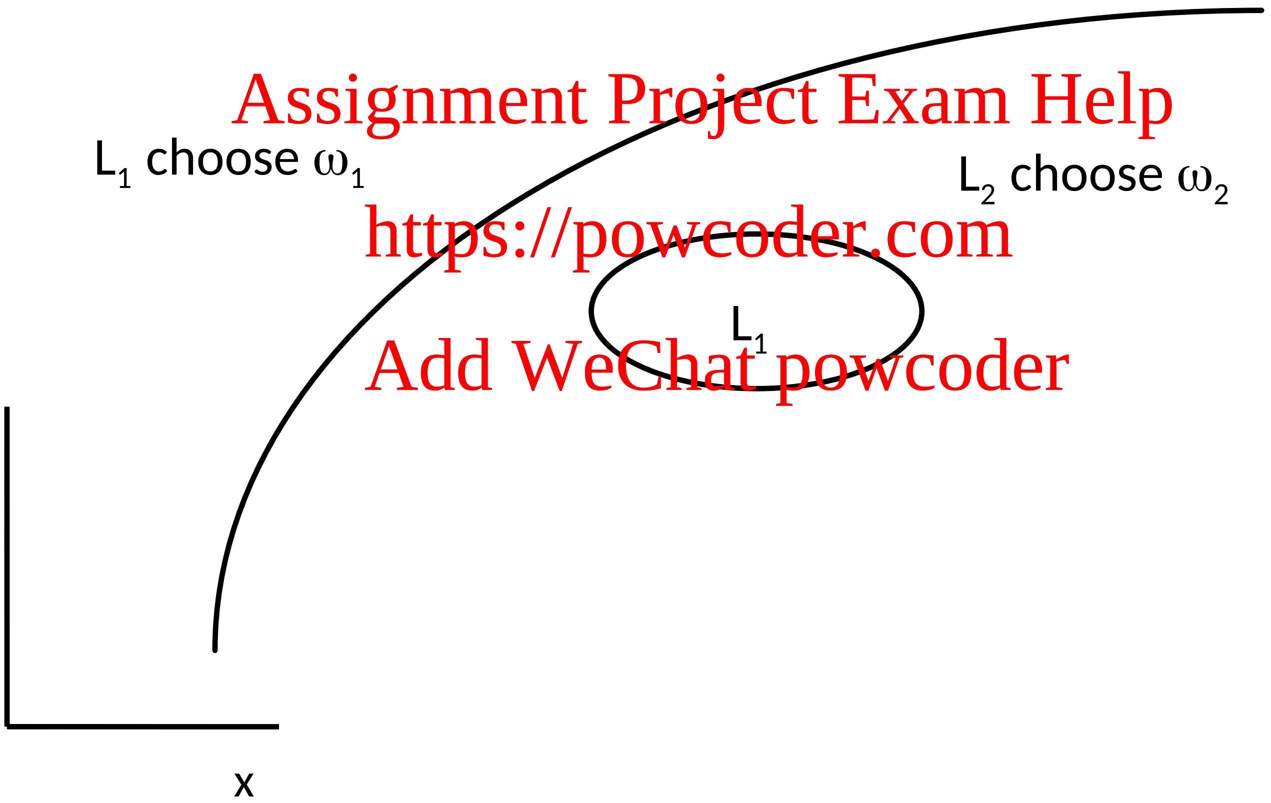
Don’t need the p(x) factor:

$$If\ \ p(x\,|\,\omega_1)\ P_1\ >\ p(x\,|\,\omega_2)\ P_2,\ \ choose\ \omega_1.\ Otherwise\ choose\ \omega_2.$$

This is, as we show next, the proper rule to use if we want to *minimize the error rate*.

Bayes Error Rate

Bayes decision rule induces a **decision surface** in the feature space --



Error Rate

Error rate for the feature vector x is

$$P(error \mid x) = \begin{cases} P(\omega_1 \mid x) & if\ choose\ \omega_2 \\ P(\omega_2 \mid x) & if\ choose\ \omega_1 \end{cases}$$

or

$$P(error \mid x) = \begin{cases} P(\omega_1 \mid x) & x \in L_2 \\ P(\omega_2 \mid x) & x \in L_1 \end{cases}$$

Minimum Error Rate

Total error rate is

$$P(error) = \int_{-\infty}^{\infty} P(error \mid x)\, p(x)\, dx$$

$$= \int_{L_1} P(\omega_2 \mid x)\, p(x)\, dx + \int_{L2} P(\omega_1 \mid x)\, p(x)\, dx$$

which is minimized if

$$P(\omega_1 | x) > P(\omega_2 | x) \quad for\ x \in L_1$$

$$P(\omega_2 | x) > P(\omega_1 | x) \quad for\ x \in L_2$$

Bayes Minimal Error Rule

- Decision rule : Assign x to the class with highest posterior

$$P(\omega_1 | x) \underset{\omega_2}{\overset{\omega_1}{\gtrless}} P(\omega_2 | x)$$

- In terms of likelihood ratios

$$l(x) \equiv \frac{p(x | \omega_1)}{p(x | \omega_2)} \underset{\omega_2}{\overset{\omega_1}{\gtrless}} \frac{P_2}{P_1} \equiv \eta \qquad (threshold)$$

- Sometimes use log likelihood ratio

$$h(x) = -\log(l(x)), \qquad h(x) \underset{\omega_1}{\overset{\omega_2}{\gtrless}} \log\left(\frac{P_1}{P_2}\right) = -\log \eta$$

Bayes Decision Surface

- e.g. Gaussian class-conditional densities

$$p(x \mid \omega_i) = \frac{1}{\sqrt{(2\pi)^N \left|\Sigma_i\right|}} \exp\left(-\tfrac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\right)$$

$$\mu_i = E[x \mid \omega_i]$$

$$\Sigma_i = \mathrm{cov}[x \mid \omega_i] = E[(x-\mu_i)(x-\mu_i)^T \mid \omega_i]$$

Likelihood ratio is

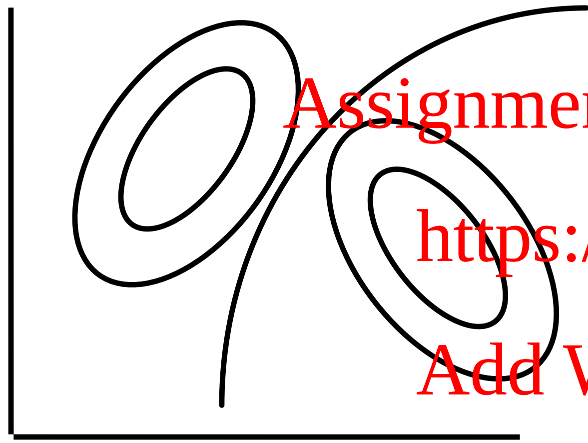
$$l(x) \equiv \frac{p(x \mid \omega_1)}{p(x \mid \omega_2)} = \sqrt{\frac{|\Sigma_2|}{|\Sigma_1|}}\ \exp -\tfrac{1}{2}\left(\left(x-\mu_1\right)^T \Sigma_1^{-1}\left(x-\mu_1\right) - \left(x-\mu_2\right)^T \Sigma_2^{-1}\left(x-\mu_2\right)\right)$$

its -log is

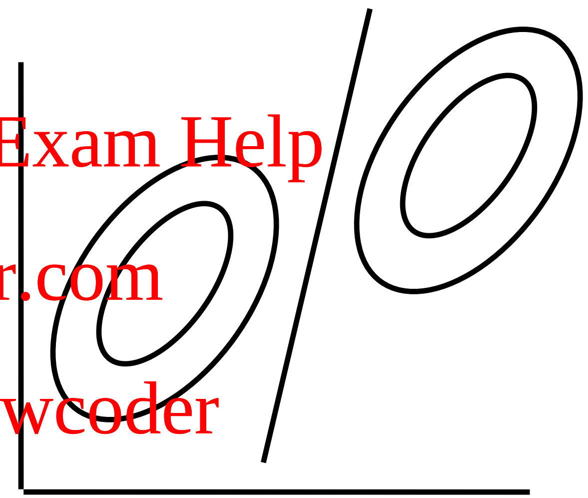
$$h(x) = -\log(l(x)) = -\tfrac{1}{2}\log|\Sigma_2| + \tfrac{1}{2}\log|\Sigma_1| + \tfrac{1}{2}\left(\left(x-\mu_1\right)^T \Sigma_1^{-1}\left(x-\mu_1\right) - \left(x-\mu_2\right)^T \Sigma_2^{-1}\left(x-\mu_2\right)\right)$$

$$= C + B^T x + x^T M x \qquad \textit{quadratic form}$$

Bayes Decision Surface
Gaussian Class Conditional Densities



$\Sigma_1 \neq \Sigma_2$



$\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$

Generalized Cost

Suppose each of the two classification error types have different cost. What's the ideal decision strategy?

1789

GEORGETOWN UNIVERSITY