METU Informatics Institute
Min720

Pattern Classification with Bio-Medical Applications

Part 9: Review

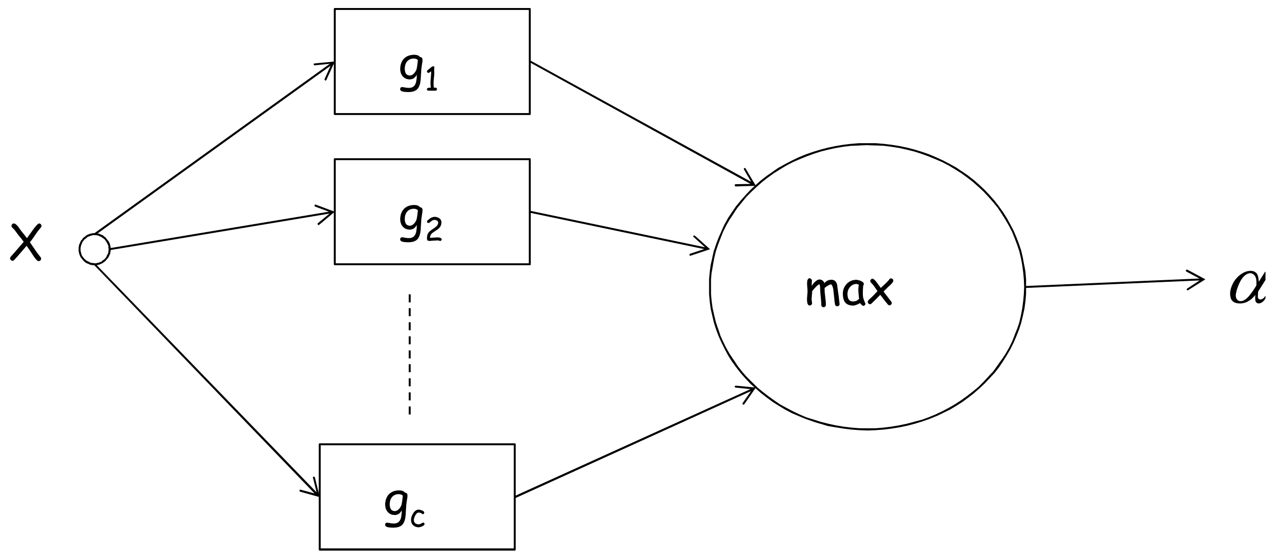
Optimum Classification : Bayes Rule

Feature vector $X = (x_1, x_2, ..., x_d)$

Categories $w_1, w_2, ..., w_c$

Discriminant functions $g_1(X), ...g_c(X)$ one per category

A pattern recognizer



Action α :

$w_1, .. w_c$ or reject

- Optimum classifier discriminant functions (minimum error/ minimum risk)

$g_i(X) = P(w_i)P(X/w_i)$ for minimum error

$P(w_i / X)$ a posteriori probability

Special Cases: Gaussian and Binary

Decision Boundaries: divide the feature space into regions $R_1, \ldots\ldots\ldots, R_c$

Gaussian= Boundaries are quadratic in general, linear for special cases.

Binary= Boundaries are linear.

Ex:

$$M_i = M \quad i = 1,2$$

$$\Sigma_i = \sigma_i^2 I$$

The decision boundary is circular.

Parameter Estimation

Given the form of $P(X/w_i)$, we want to estimate its parameters using the samples $X_1,......X_n$ from a given category.

Maximum Likelihood Estimation:

Maximize the likelihood function

$$L = \prod_{i=1}^{n} P(X_i;\theta)$$

Find θ that will cause $\nabla L = 0$

$$\nabla L = \left[\frac{\partial L}{\partial \theta_1}, \frac{\partial L}{\partial \theta_2}, ... \right]^T \quad \text{for} \quad \theta = (\theta_1, \theta_2, ...\theta_n)$$

Ex: if $P(x/\theta) = \theta^2 x e^{-\theta x} \quad for \quad x > 0$

Then,

$$\log P = +2\log\theta + (-\theta X) + \log X$$

$$\log L = \log \prod p(X_i;\theta) = \sum \log P(X_i, \theta)$$

$$\frac{dL}{d\theta} = +\frac{2n}{\theta} - \sum_{i=1}^{n} X_i = 0 \Rightarrow \theta = \frac{2n}{\sum X_i} = \frac{2}{\overline{X}}$$

Nearest Neighbor Classification

- Density Estimation
- 1-NN, k-NN
- Shown to approach Bayes Rate in error, but computationaly heavy burden.

Reduce computational problems with editing algorithms.

Reducing the number of features

-to remove the redundancies and obtain statistically independent features: use PCA (Principle Component Analysis)

- To obtain features with good separating ability- Fisher's linear discriminant

Linear and Generalized Discriminant Functions

Decision Boundary : Linear

How to find the parameters of hyperplanes that separate best.

- Iterative methods: Use gradient descent: Perceptron learning – minimizes the misclassified samples
- Non-iterative methods- minimum squared error (Widrow-Hoff)
- Support vector machines- Dimension is increased so that samples are separable.

Generalized Discriminant Functions:

-Functions like quadratic may be used to generalize.

- Multi-category problems.

Example:

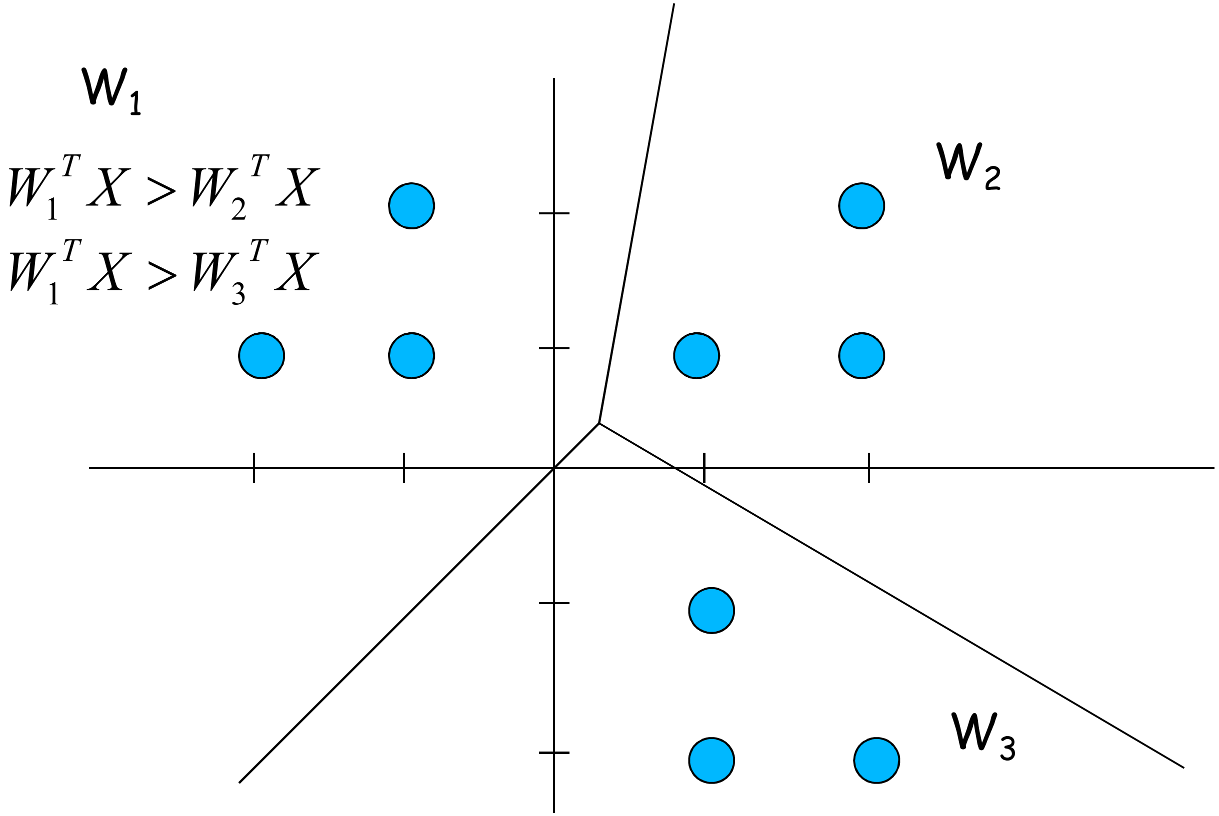
$$c_1 : (1,1)^T, (2,2)^T, (2,1)^T$$

$$c_2 : (1,-1)^T, (1,-2)^T, (2,-2)^T$$

$$c_3 : (-1,1)^T, (-1,2)^T, (-2,1)^T$$

a) Is the problem linearly separable?

b) Obtain 3 weight vectors using perceptron algorithm



To solve, apply the algorithm with multicategory extension.

We need to W_{10}, W_{20}, W_{30}

$$W_{10} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$
$$W_{20} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \quad \text{augmented}$$
$$W_{30} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

Augment the samples:

$$C_1 = [1 \quad 1 \quad 1]^T, [2 \quad 2 \quad 1]^T \ldots\ldots$$

$$C_2 = \ldots\ldots\ldots similarly$$

Iterate the algorithm starting with a randomly selected sample.

$$W_{11} = W_{10} + Y_1 \quad \left(Y_1 = \begin{matrix} 1 \\ 1 \\ 1 \end{matrix} \right)$$

if $W_{10} Y_1 \leq 0$

so $W_{11} = Y_1$