Pattern Recognition

Introduction to Probability

Bayes Decision Theory

Topics

- Definitions (Review)
- Classification Based on Bayes Decision Theory
- Probability Theory
- Bayes' Theorem

Review

- Features: These are measurable quantities obtained from the patterns, and the classification task is based on their respective values.

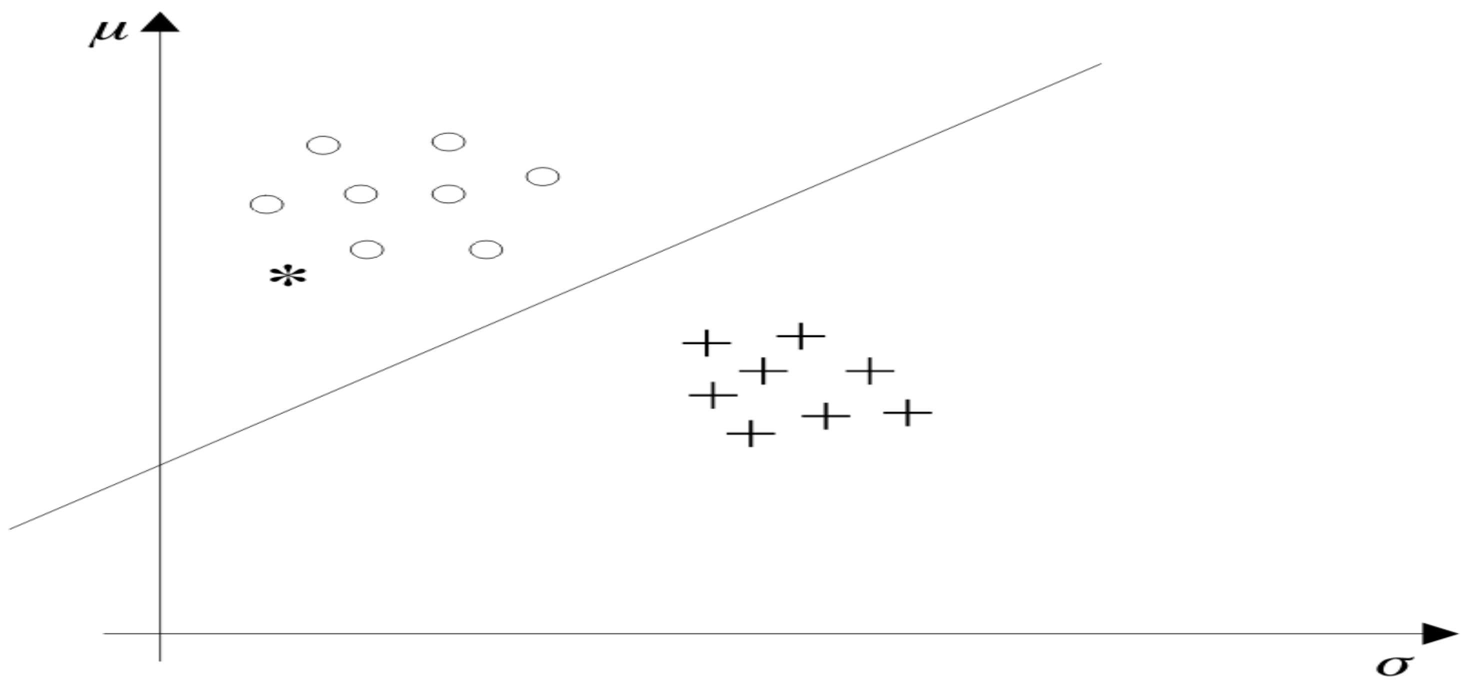
- Feature vectors: A number of feature values $<x_1, x_2, .., x_n>$ constitute the feature vector

- Feature vectors are treated as random vectors.

- A random variable is a variable whose value is subject to variations due to chance and it can take on a set of possible different values, each with an associated probability.

Review

- The classifier consists of a set of functions, whose values, computed at $\underline{X}$ determine the class to which the corresponding pattern belongs



Classification Based on Bayes Decision Theory

- Statistical nature of feature vectors

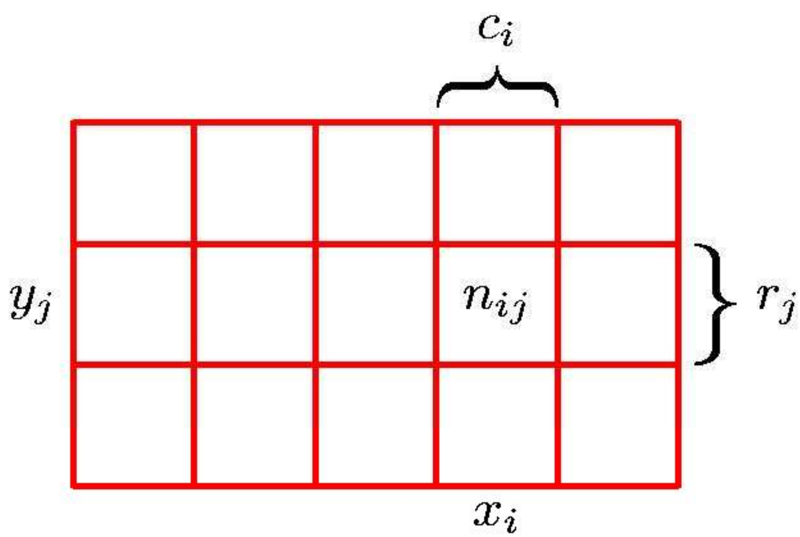
$$\underline{x} = [x_1, x_2, \ldots, x_l]^T$$

- Assign the pattern represented by feature vector $\underline{x}$ to the most probable of the available classes

$$\omega_1, \omega_2, \ldots, \omega_M$$

That is $\underline{x} \rightarrow \omega_i : P(\omega_i | \underline{x})$ maximum

Probability Theory



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

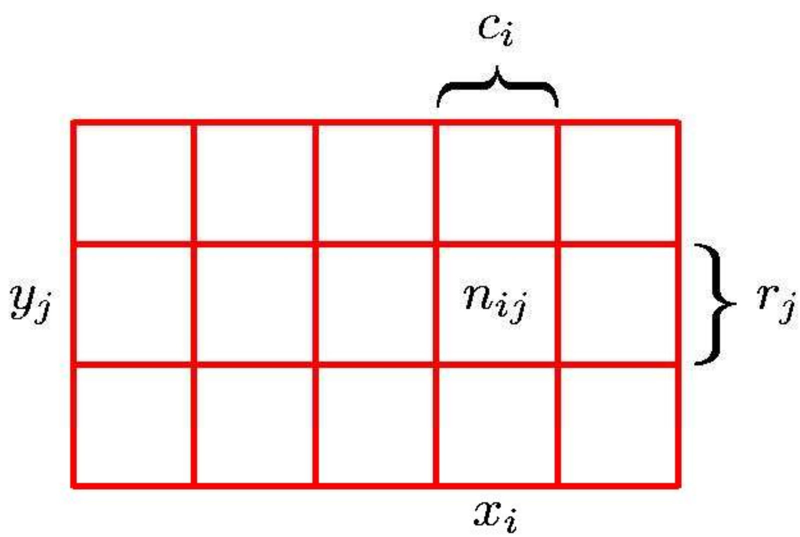
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Probability Theory



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$
$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$
$$= p(Y = y_j | X = x_i) p(X = x_i)$$

The Rules of Probability

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

Example

- A jar contains 10 black and 10 white marbles. Two marbles are chosen without replacement. What is the probability of drawing a white marble and then a black one?

$P(x=\text{white}, y=\text{black}) = P(y|x).P(x) = (10/19)(1/2)$

Example

- A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34, and the probability of selecting a black marble on the first draw is 0.47. What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?
- $P(x=black, y=white)=0.34=p(y|x).p(x)$
- $P(y|x)=0.72$

Independent probability

- If two events, A and B are independent then the joint probability is

$$P(A \text{ and } B) = P(A) \, . \, P(B)$$

a-posteriori probabilities

- The a-posteriori probability is the probability of the parameter x given the evidence T and is shown as : p(x|T)

- a-priori probability of an uncertain quantity *x* is the probability distribution that would express one's uncertainty about *x* before some evidence is taken into account.

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_{Y} p(X|Y)p(Y)$$

Example

- Suppose there is a box of marbles with 60% of them being square and 40% being round. The round marbles are either black or white in equal numbers; the square marbles are all black. An observer sees a (random) marble from a distance; all the observer can see is that this marble is black. What is the probability this marble is round?
- P(round)=0.4
- P(black|round)=0.5
- P(black)=p(black|round).p(round) + p(black|square).p(square) = 0.5*0.4+1*0.6 = 0.8
- P(round|black)=(p(black|round).p(round))/P(black)
- 0.5x0.4/0.8=0.25

The Bayes' classification rule (for two classes $M=2$)

- Given $\underline{x}$ classify it according to the rule

$$\begin{aligned} &If\ P(\omega_1|\underline{x}) > P(\omega_2|\underline{x})\ \ \underline{x} \rightarrow \omega_1 \\ &If\ P(\omega_2|\underline{x}) > P(\omega_1|\underline{x})\ \ \underline{x} \rightarrow \omega_2 \end{aligned}$$

- Equivalently: classify $\underline{x}$ according to the rule

$$p(\underline{x}|\omega_1)P(\omega_1)(><)p(\underline{x}|\omega_2)P(\omega_2)$$

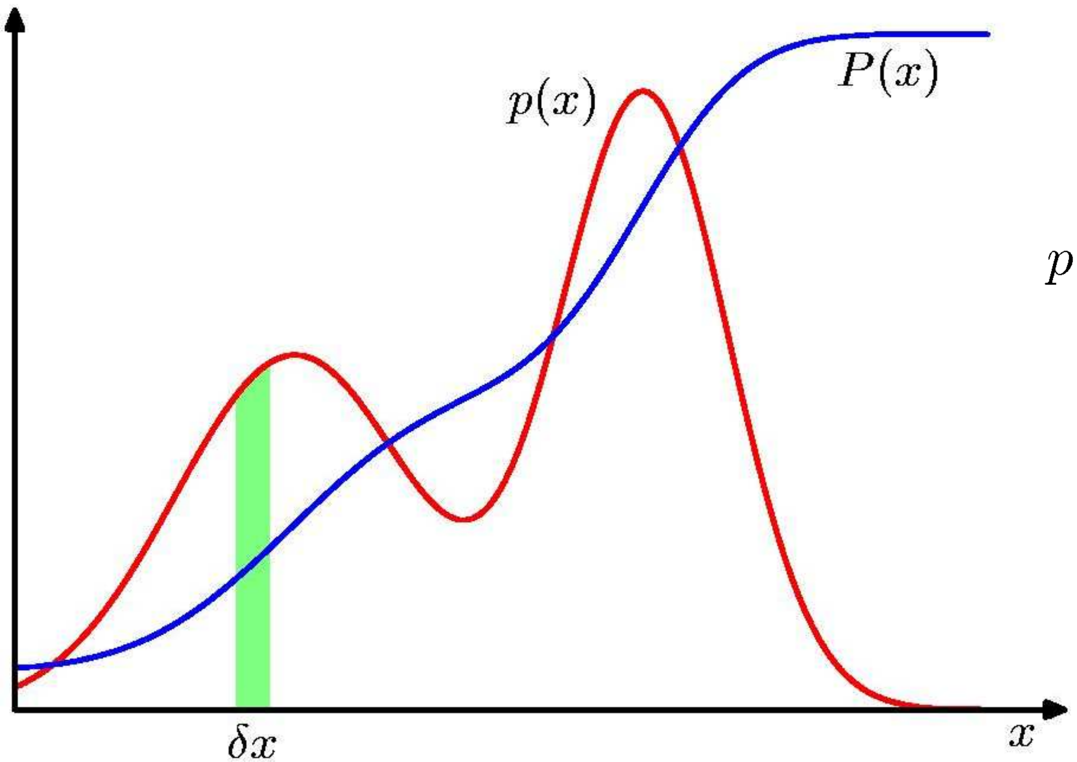
- For equi-probable classes the test becomes

$$p(\underline{x}|\omega_1)(><)P(\underline{x}|\omega_2)$$

Probability Density Function

- A **probability density function** (**pdf**), or **density** of a continuous random variable, is a function that describes the relative likelihood for this random variable to take on a given value.

Probability Densities



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

$$p(x) \geqslant 0 \qquad \int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

Cumulative Distribution Function

- **Cumulative distribution function** (**CDF**), or just **distribution function**, describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to x.

Expectations

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

$$\mathbb{E}[f] = \int p(x) f(x)\, \mathrm{d}x$$

$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation
(discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

Approximate Expectation
(discrete and continuous)

Variances and Covariances

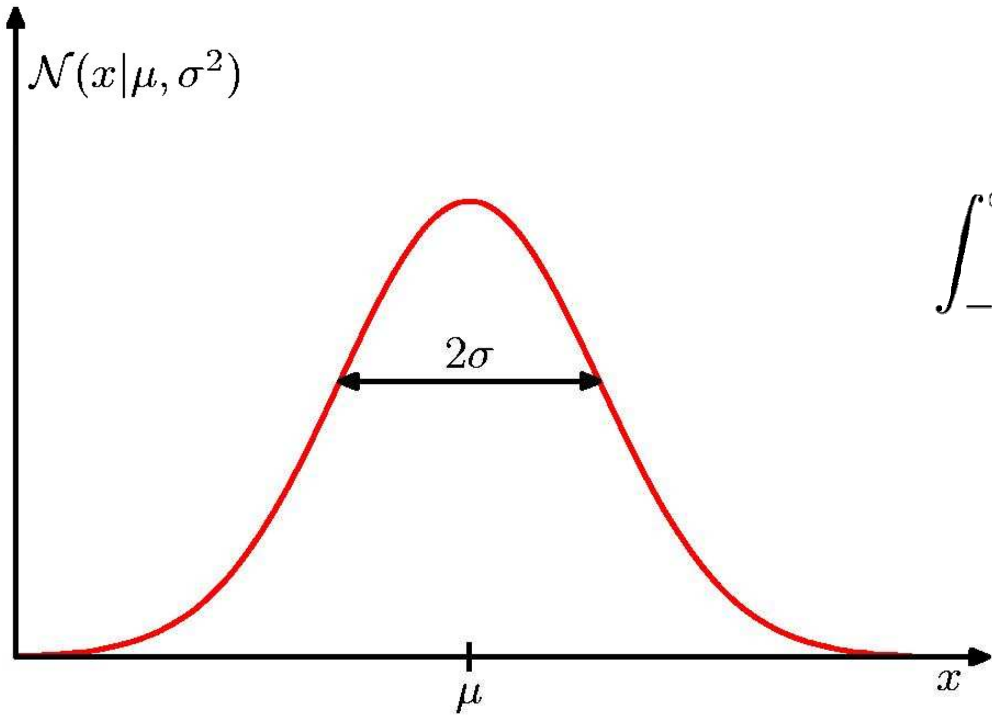
$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$$\begin{aligned}\mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

$$\begin{aligned}\mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}]\end{aligned}$$

The Gaussian Distribution

$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(x|\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right)\,\mathrm{d}x = 1$$

Gaussian Mean and Variance

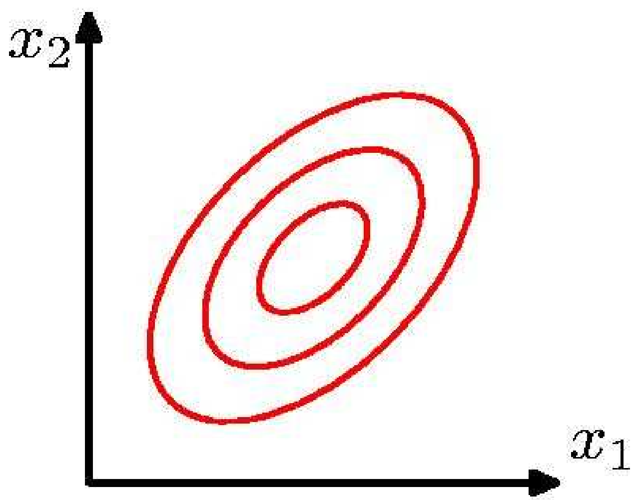
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \, \mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x^2 \, \mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

The Multivariate Gaussian

$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$



Questions?