Pattern Recognition

Introduction to Probability

Bayes Decision Theory (2)

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

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

$$
\begin{array}{l}
If \;\; P(\omega_1 | \underline{x}) > P(\omega_2 | \underline{x}) \;\; \underline{x} \rightarrow \omega_1 \\
If \;\; P(\omega_2 | \underline{x}) > P(\omega_1 | \underline{x}) \;\; \underline{x} \rightarrow \omega_2
\end{array}
$$

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

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

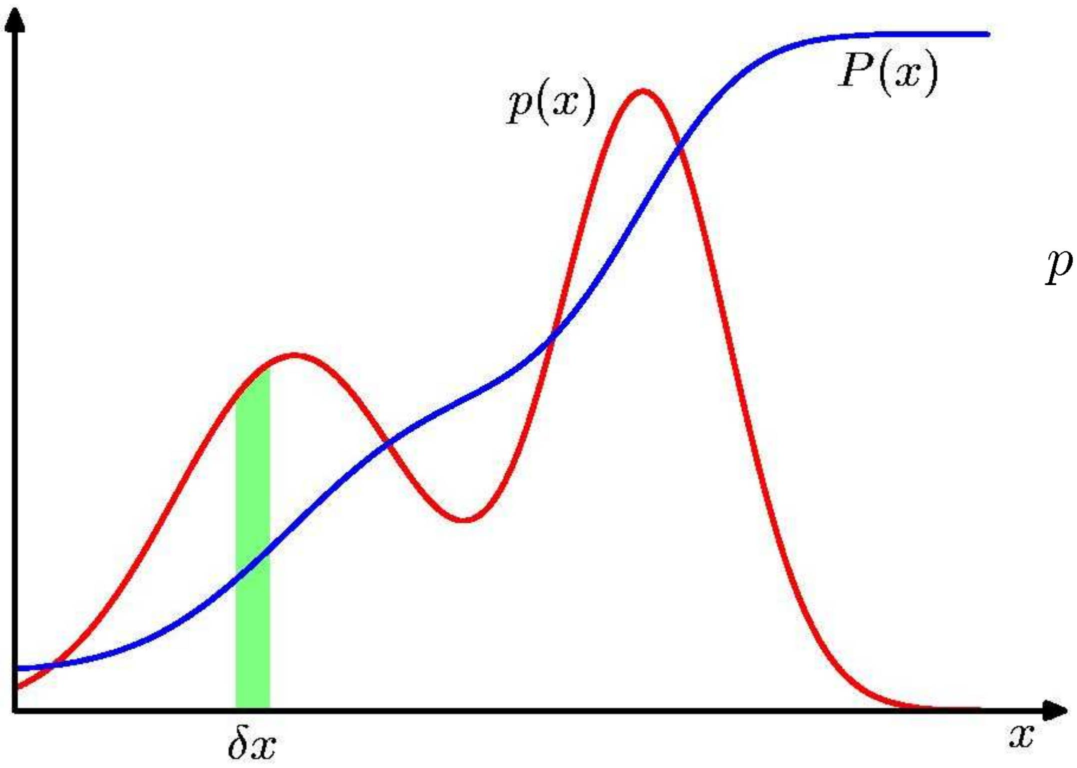
- For equiprobable classes the test becomes

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

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

- The variance (σ^2) is a measure of how far each value in the data set is from the mean.
- Covariance is a measure of how much two random variables change together.

Variances and Covariances

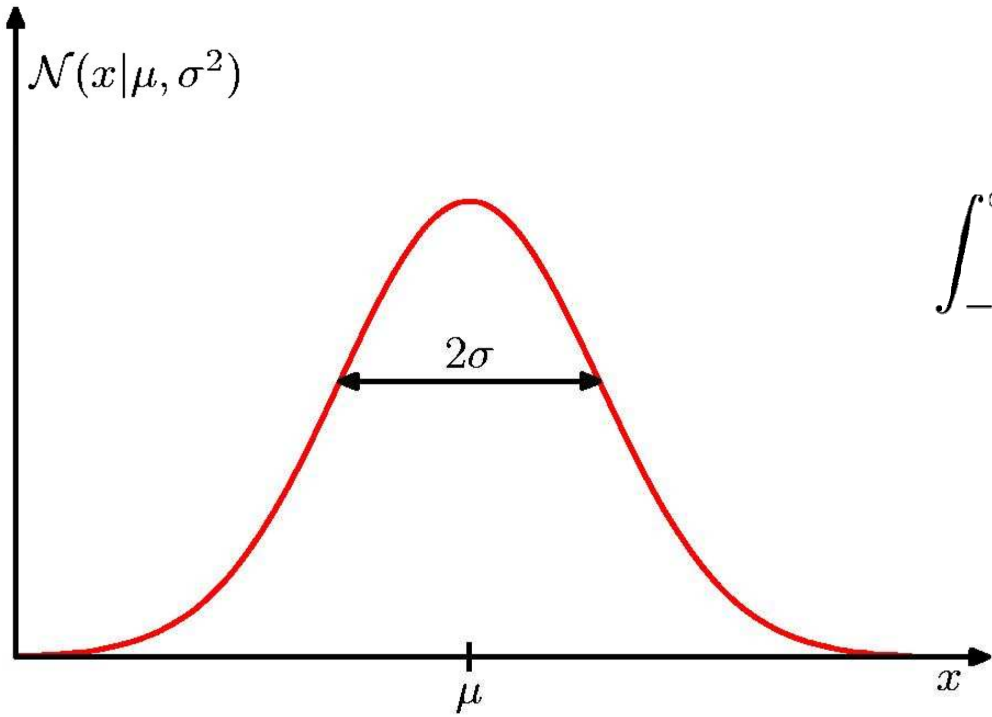
$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

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

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

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

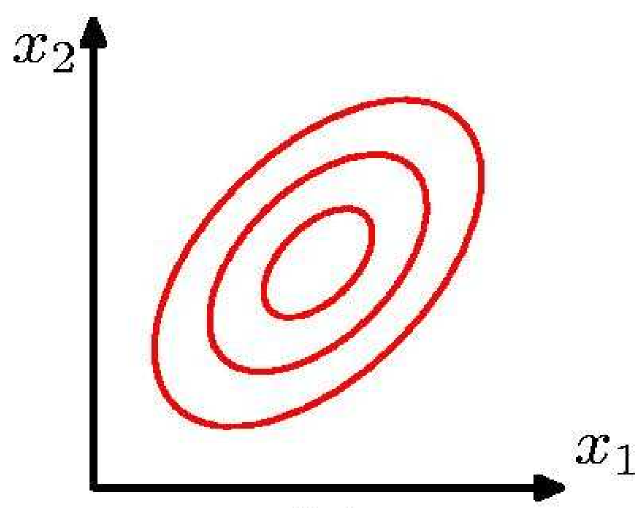
$$\mathbb{E}[x] = \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) x \,\mathrm{d}x = \mu$$

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

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

The Multivariate Gaussian

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



Classification Problem

- Feature vectors x that we aim to classify belong to the feature space *F*.
- The task is to assign an arbitrary feature vector $x \in F$ to one of the *c* classes
- We know the
  - 1. prior probabilities $P(\omega_1), \ldots, P(\omega_c)$ of the classes and
  - 2. the class conditional probability density functions $p(x \mid \omega_1), \ldots, p(x \mid \omega_c)$.

Classification Problem

- The classification problem is to assign an arbitrary feature vector $x \in F$ to one of c classes.
- The classifier is a function from the feature space onto the set of classes,
  $\alpha : F \rightarrow \{ \omega_1 , . . . , \omega_c \}$.   ($\alpha(x)$ is the classifier)
- The parts of the feature space corresponding to classes $\omega_1 , . . . , \omega_c$ are denoted by $R_1 , . . . , R_c$

Classification Error

- Objects having same feature vectors can belong to different classes.
- For a specified classification problem our task to derive classifier that makes as few errors (misclassifications) as possible.
- Classification error by the classifier α is denoted by $E(\alpha)$.

Classification Error

$E(\alpha) = 1 - P\,(\alpha(x)\ |\ x)$

$$E(\alpha) = \int_{\mathbb{F}} [1 - P(\alpha(\mathbf{x})|\mathbf{x})] p(\mathbf{x}) d\mathbf{x}$$

$$P(\alpha(\mathbf{x})|\mathbf{x}) = \frac{p(\mathbf{x}|\alpha(\mathbf{x}))P(\alpha(\mathbf{x}))}{p(\mathbf{x})}.$$

$$E(\alpha) = 1 - \int_{\mathbb{F}} p(\mathbf{x}|\alpha(\mathbf{x}))P(\alpha(\mathbf{x})) d\mathbf{x},$$

Bayes Minimum Error Classifier

- The Bayes classifier is defined as

$$\alpha_{Bayes}(x) = \arg \max_{\omega_i, i=1,\ldots,c} P(\omega_i \mid x).$$

the Bayes classifier selects the most probable class when the observed feature vector is x

Bayes Minimum Error Classifier

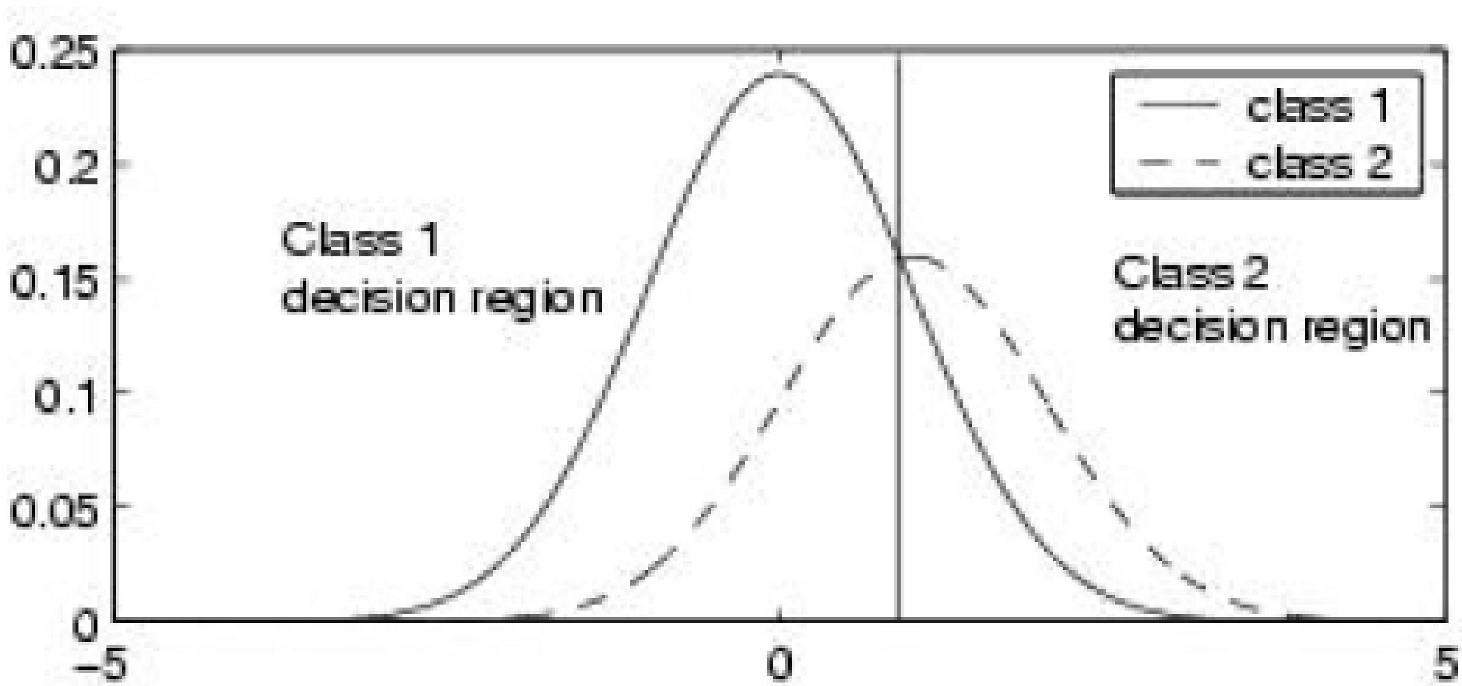
- Using the posterior probability

$$\alpha_{Bayes}(x) = \arg \max_{\omega_i, i=1,\dots,c} p(x \mid \omega_i) P(\omega_i)$$

That is to say: the Bayes classifier minimizes the conditional error $E(\alpha(x) \mid x) = 1 - P(\alpha(x) \mid x)$

for all x

Classification Error



Bayes Minimum Risk Classifier

- The Bayes minimum error classifier is a special case of the more general Bayes minimum risk classifier.
- Given actions $\alpha_1 , . . . , \alpha_a$
- The actions are tied to the classes via a loss function λ
- The value $\lambda(\alpha\ i \mid \omega\ j\ )$ quantifies the loss incurred by taking an action α_i when the true class is ω_j .

Bayes Minimum Risk Classifier

- The Bayes minimum risk classifier chooses the action with the minimum conditional risk.
- The conditional risk of taking the action α_i when the observed feature vector is x, is defined as

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j) P(\omega_j|\mathbf{x}).$$

Assignment

- Assume a classifier should identify handwritten numbers
- Assume each number is given as a 16x8 binary matrix
- Choose a set of features (feature space)
  - E.g. total number of 1s (black points), location of the column with max 1s, location of rows with max 1s, variance for columns/rows, etc.

Assignment

- For each number create 5 samples (at least)
- Draw pdf for one of the numbers
- Compute covariance
- Develop classifier
- Use some sample input (other than the ones used for training) and classify them
- Discuss the success/failure rates of your classifier