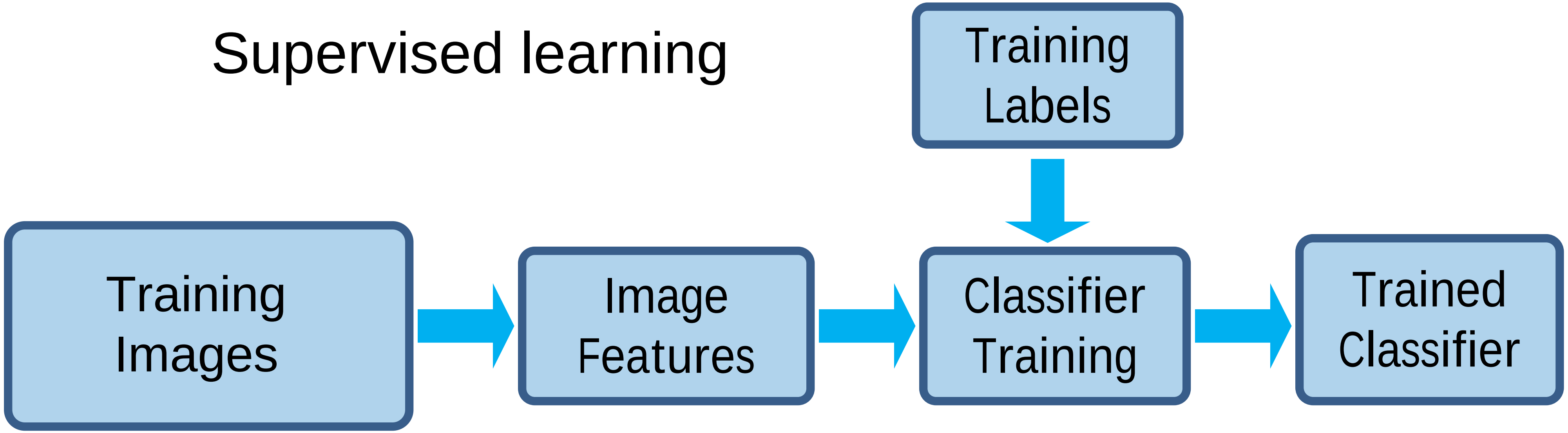
Lecture 9.3
Introduction to Machine Learning

Idar Dyrdal

Machine learning (Pattern recognition)

- Recognition of individuals (instance recognition)
- Discrimination between classes (pattern recognition, classification)

Supervised learning



Pattern recognition in practice

Working applications of Image Pattern recognition:

- Reading license plates, postal codes, bar codes
- Character recognition
- Automatic diagnosis of medical samples
- Fingerprint recognition
- Face detection and recognition
- …

Classification system

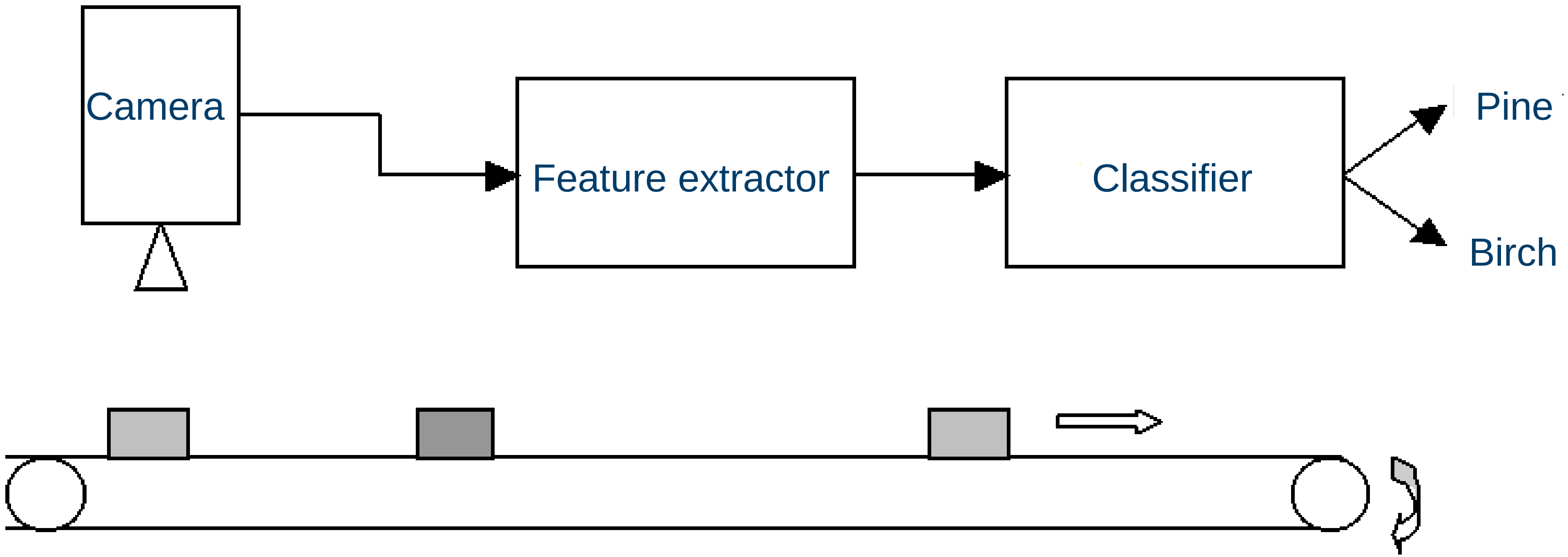
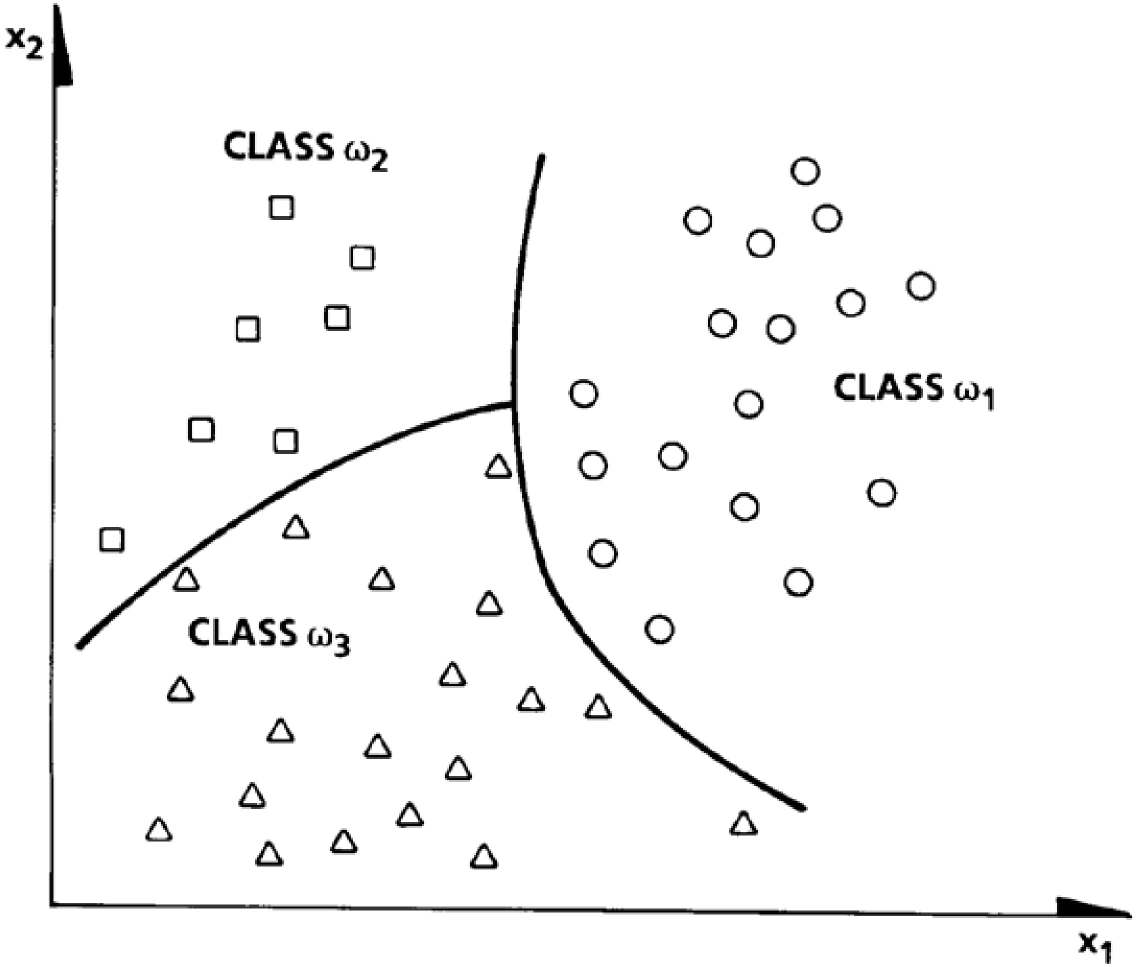


Image features for object recognition

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_d \end{bmatrix}$$

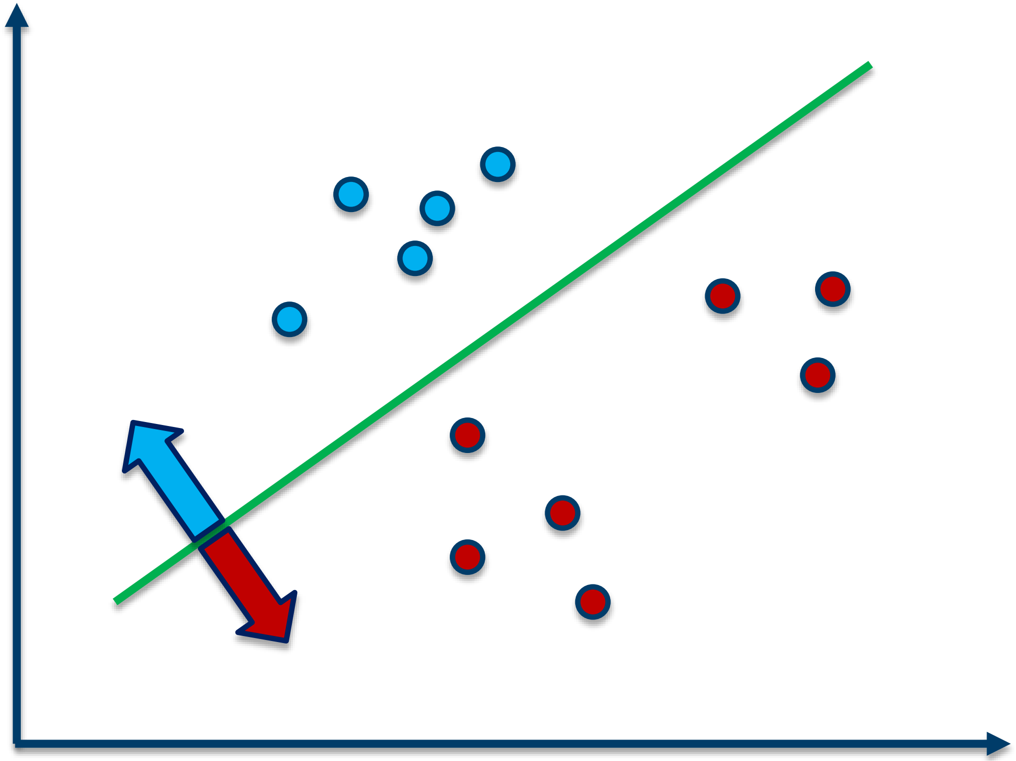
Feature vector and feature space

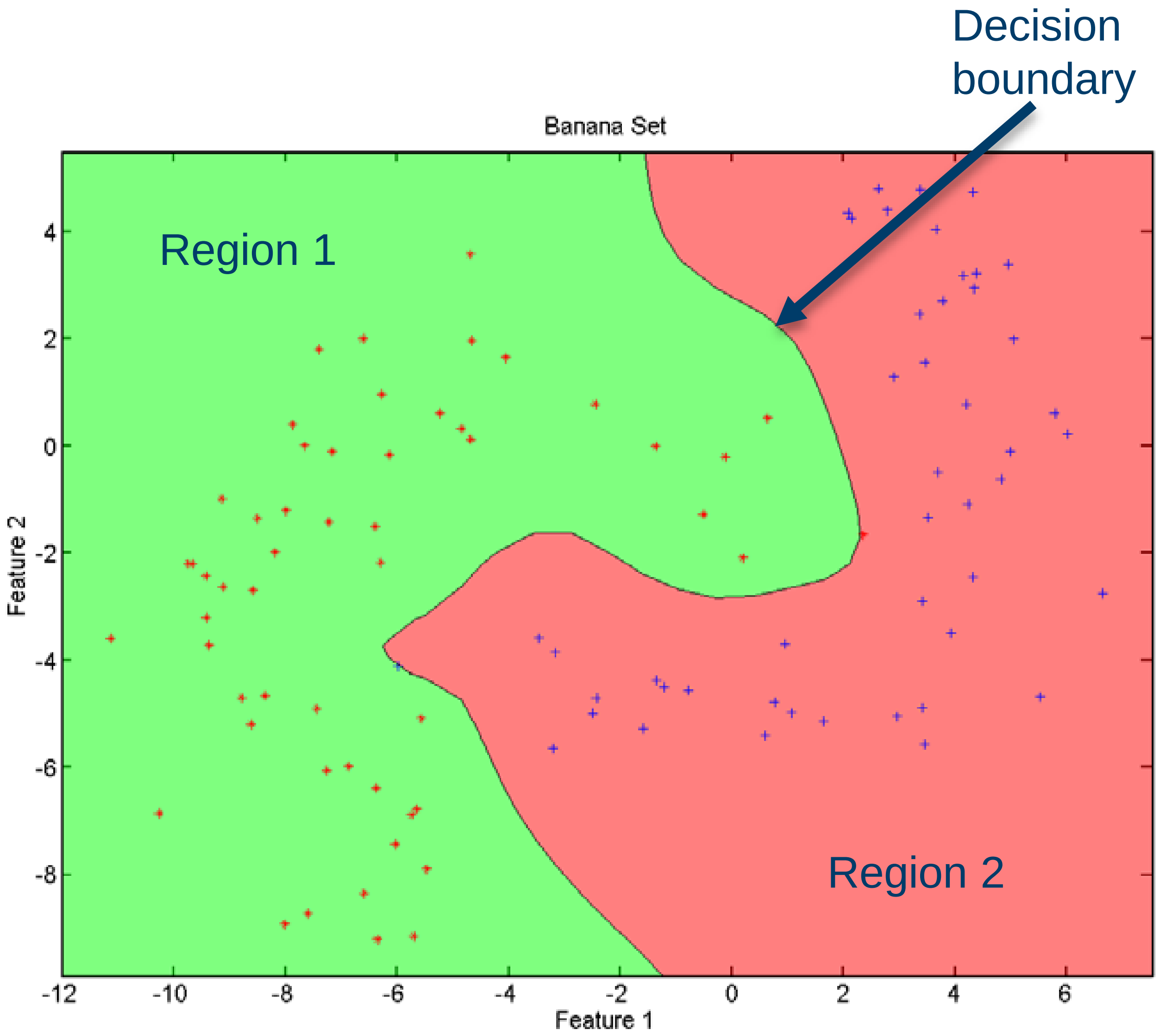
$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_d \end{bmatrix}$$



Training of classifiers

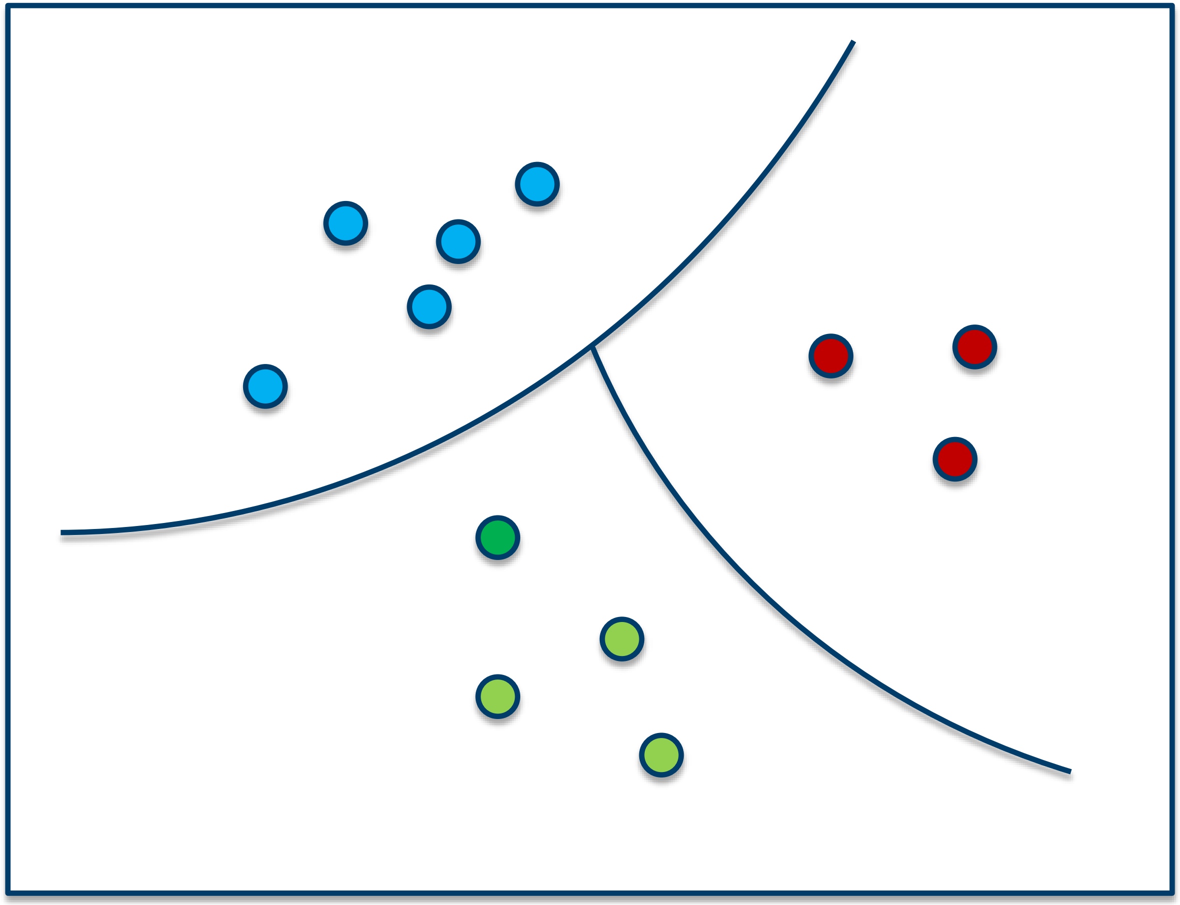
Learn a function to predict the class from the given features



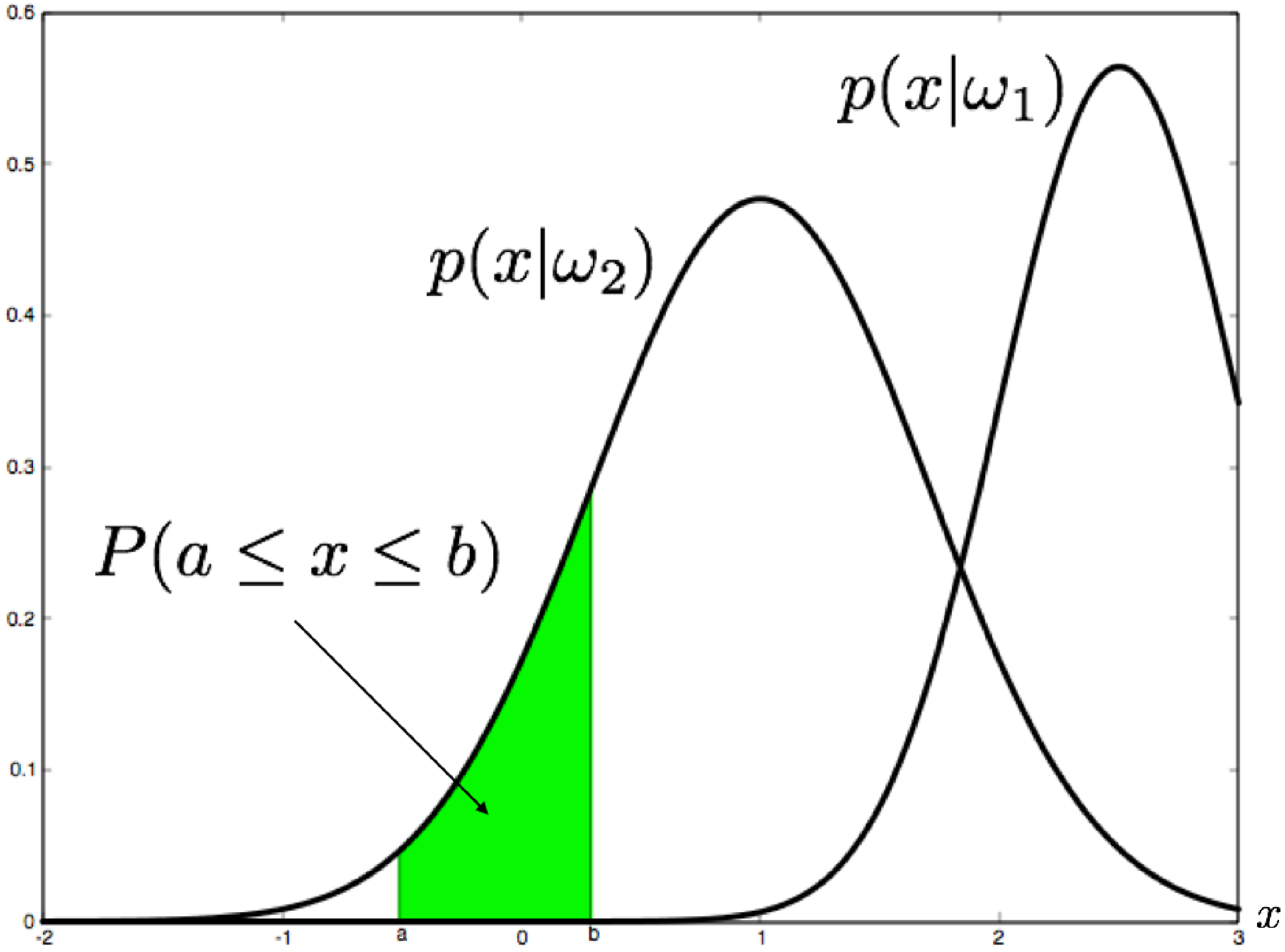


Classifiers and training methods

- Bayes classifier
- Nearest-neighbors and K-nearest-neighbors
- Parzen windows
- Linear and higher order discriminant functions
- Neural nets
- Support Vector Machines (SVM)
- …



Class conditional probability density functions



Bayesian decision theory

Overview

Class conditional densities:

$$p(\vec{x}|\omega_i), \text{ for each class } \omega_1, \omega_2, \ldots, \omega_c$$

Prior probabilities:

$$P(\omega_1), P(\omega_2), \ldots, P(\omega_c)$$

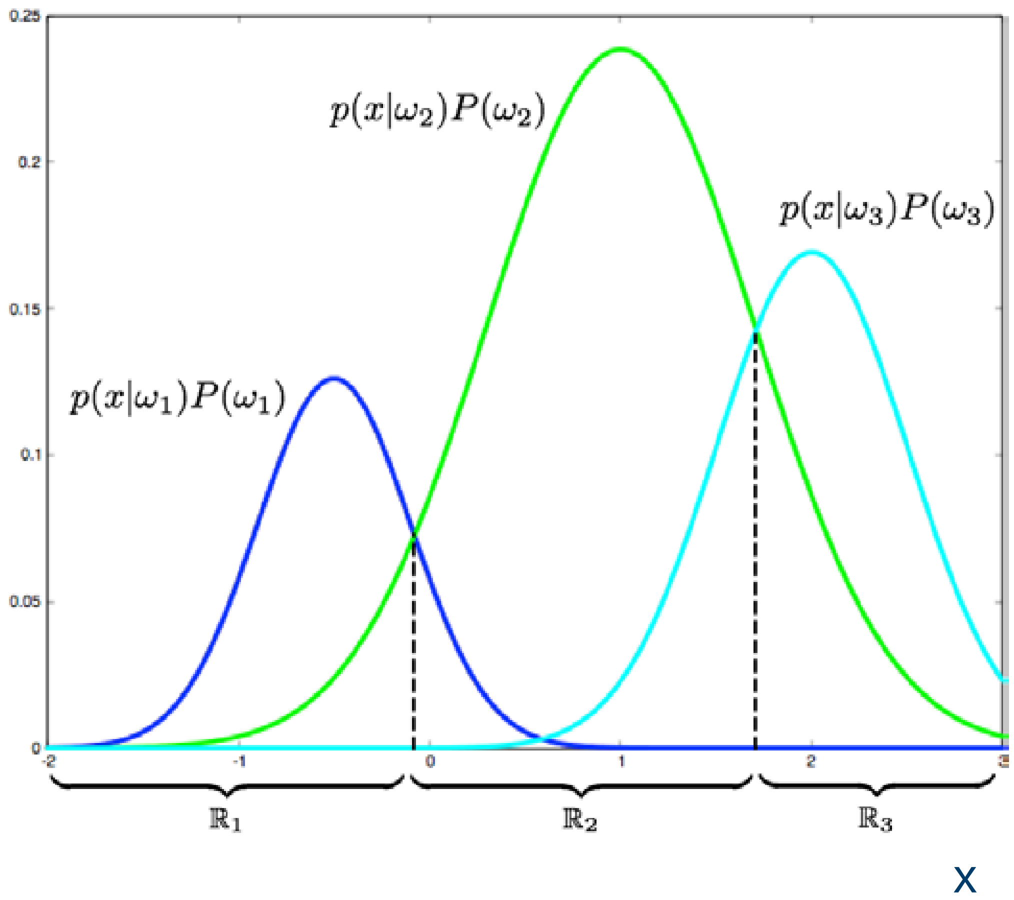
Posterior probabilities given by Bayes rule:

$$P(\omega_i|\vec{x}) = \frac{p(\vec{x}|\omega_i)P(\omega_i)}{\sum_{j=1}^{c} p(\vec{x}|\omega_j)P(\omega_j)}, \ i = 1, \ldots, c$$

(a function of the measured feature vector $\vec{x} = [x_1, x_2, \ldots, x_d]^t$).

Minimum error rate classification:

Assign the unknown object to the class with maximum posterior probability!



Density estimation

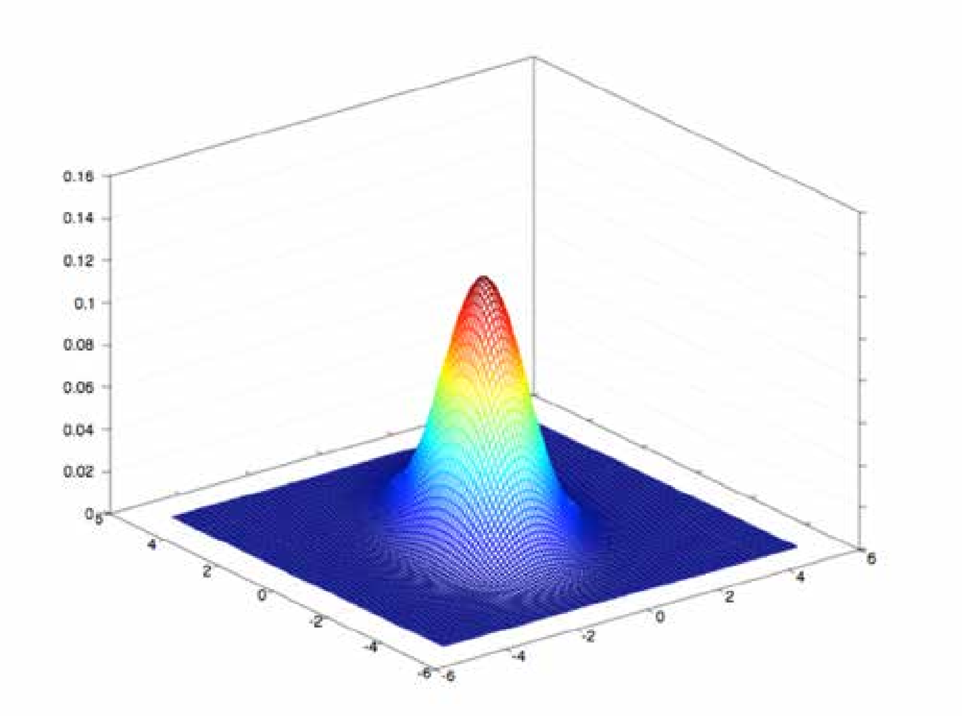
Parametric methods:

- Assume a given shape of the density function
- Use the training set to estimate the unknown parameters.

Non-parametric (distribution free) methods:

- Point estimation of the density using the training set directly
- Parzen windows
- Nearest neighbor estimation (leads directly to the nearest-neighbor and k-nearest-neighbor classifiers).

Example – Gaussian distribution:

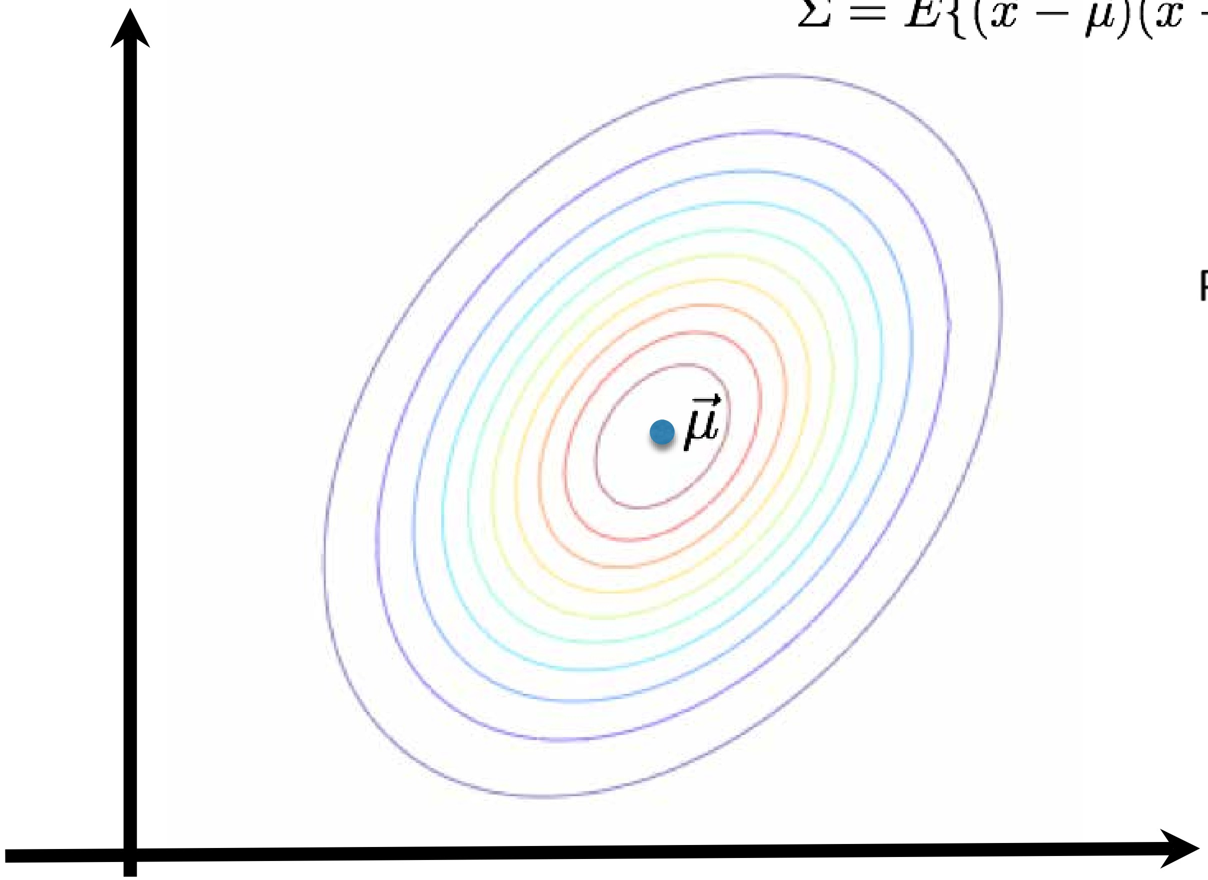


$$p(\vec{x} \mid \omega_i) = \frac{1}{(2\pi)^{d/2} \mid \Sigma_i \mid^{1/2}} \exp[-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^t \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)]$$

Parameters: $\vec{\mu}_i$ and Σ_i

Parameter estimation

$$\Sigma = E\{(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^t\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1d} \\ \vdots & \vdots & & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_{dd} \end{bmatrix}$$



Parameter estimates:

$$\hat{\vec{\mu}}_i = \vec{m}_i = \frac{1}{n}\sum_{k=1}^{n} \vec{x}_k$$

$$\hat{\Sigma}_i = \frac{1}{n}\sum_{k=1}^{n} (\vec{x}_k - \vec{m}_i)(\vec{x}_k - \vec{m}_i)^t$$

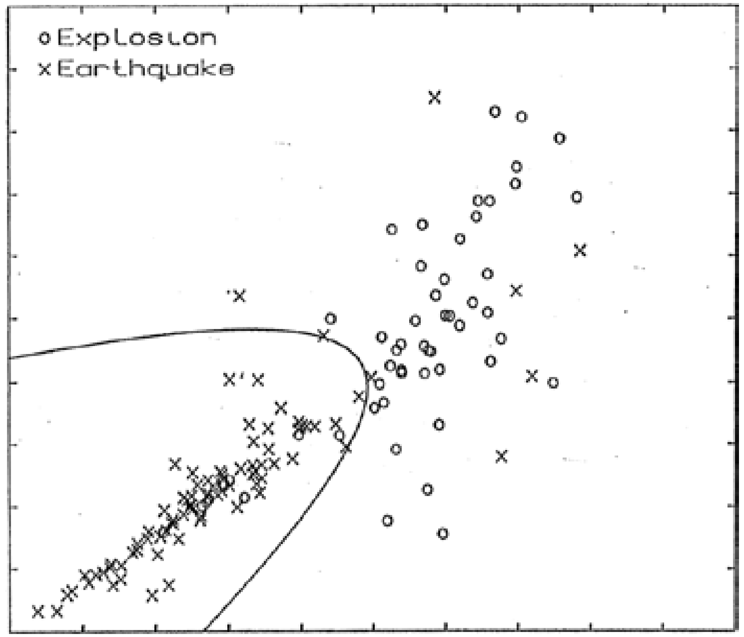
Discriminant functions

Estimate of the density in a given point:

$$\hat{p}(\vec{x} \mid \omega_i) = \frac{1}{(2\pi)^{d/2} \mid \hat{\Sigma}_i \mid^{1/2}} \exp[-\frac{1}{2}(\vec{x} - \hat{\vec{\mu}}_i)^t \hat{\Sigma}_i^{-1} (\vec{x} - \hat{\vec{\mu}}_i)]$$

From Bayes rule:

$$\hat{P}(\omega_i \mid \vec{x}) = \frac{\hat{p}(\vec{x} \mid \omega_i) P(\omega_i)}{\sum_{j=1}^{c} \hat{p}(\vec{x} \mid \omega_j) P(\omega_j)}$$



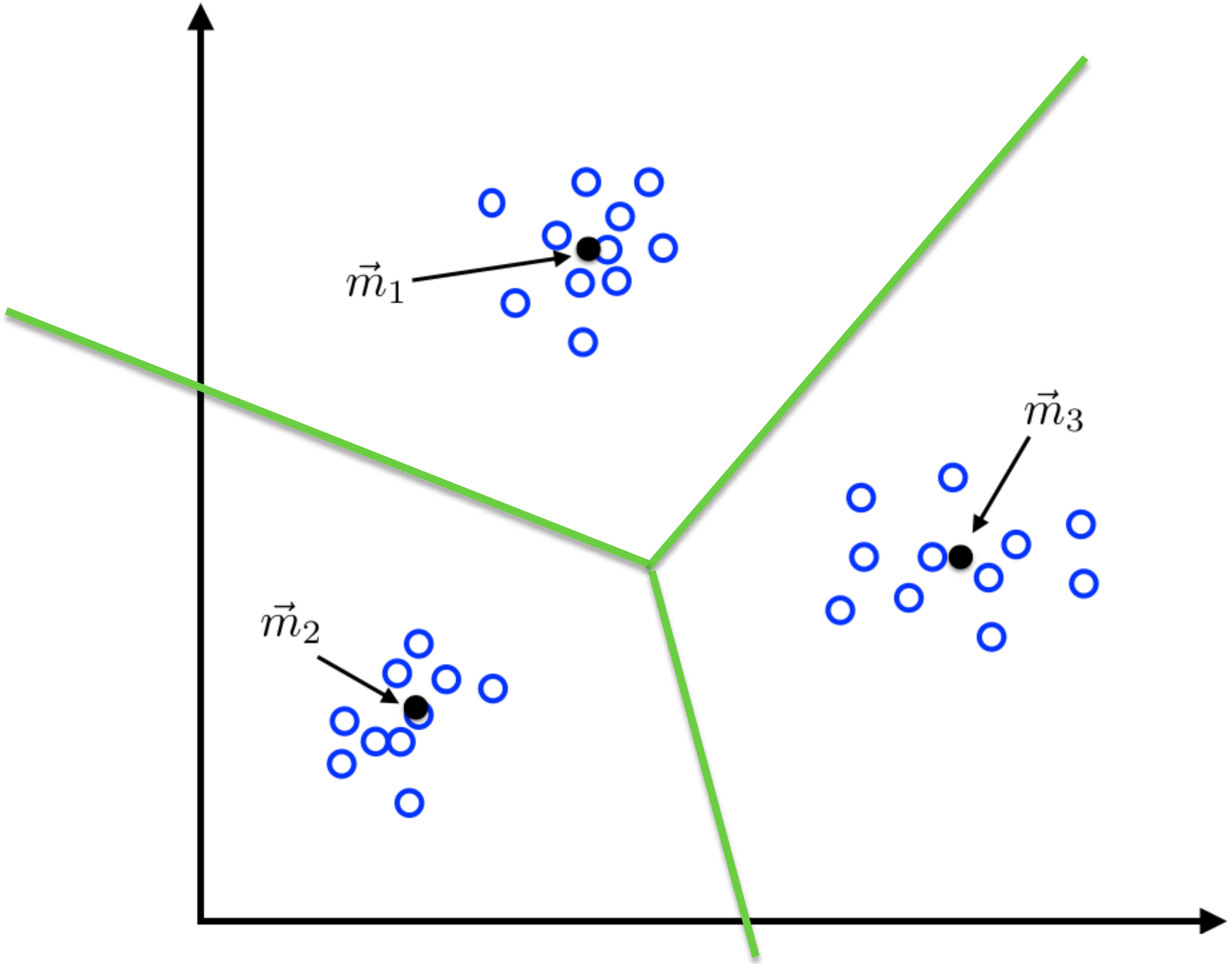
Example of a discriminant function:

$$g_i(\vec{x}) = \ln \hat{P}(\omega_i \mid \vec{x}) = \ln \hat{p}(\vec{x} \mid \omega_i) + \ln P(\omega_i)$$

Decision rule:

Choose the class with maximum discriminant function value.

Example - linear classifier



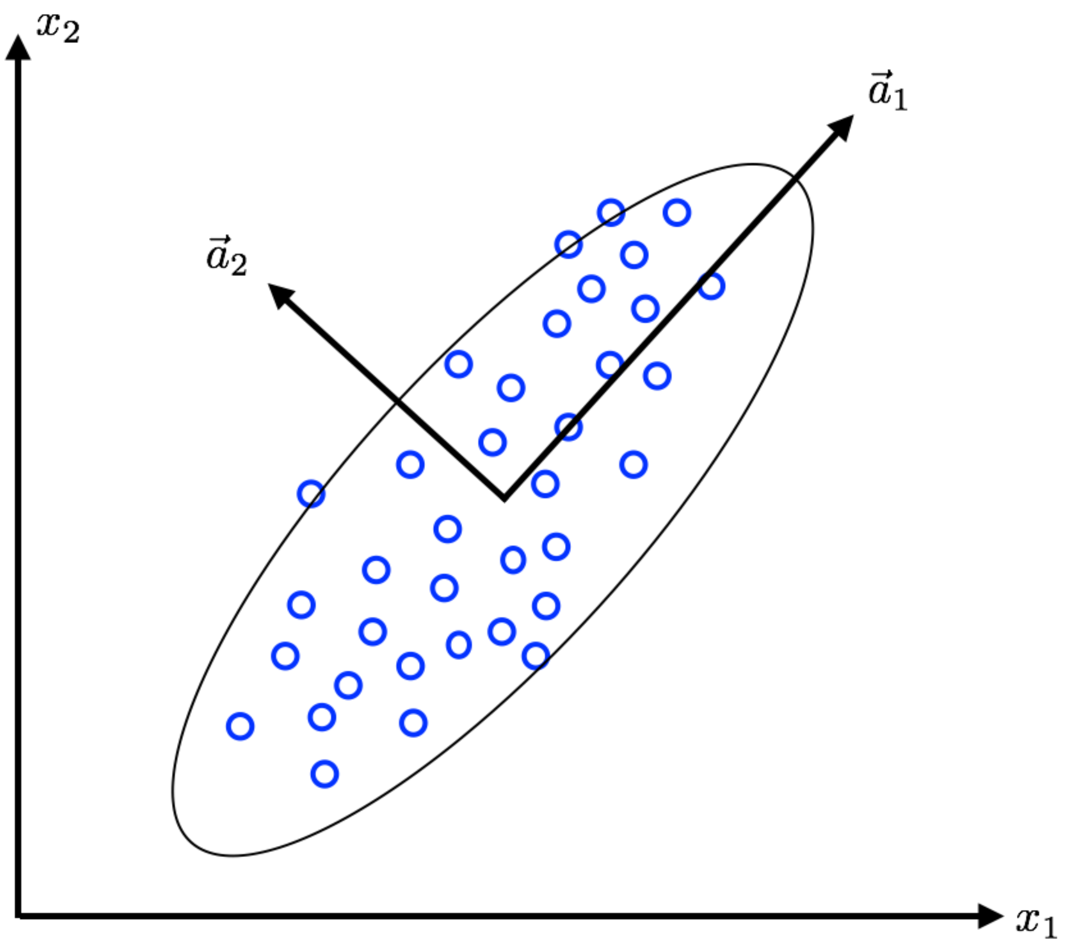
Uncorrelated features and common covariance matrices

$\Downarrow$

Linear decision boundaries

Dimensionality reduction - linear transformations

- **PCA**, ICA, LLE, Isomap
- PCA (Principal Components Analysis) is one of the most important techniques for dimensionality reduction
- It takes advantage of correlations between the features to produce the best possible lower dimensional representation of the data with respect to reconstruction error
- The eigenvectors of the lumped covariance matrix defines the new features in the transformed feature space.



Summary

Recognition:

- Pattern classification
- Training of classifiers (supervised learning)
- Parametric and non-parametric methods
- Discriminant functions
- Dimensionality reduction

Read also: Szeliski 14.1