

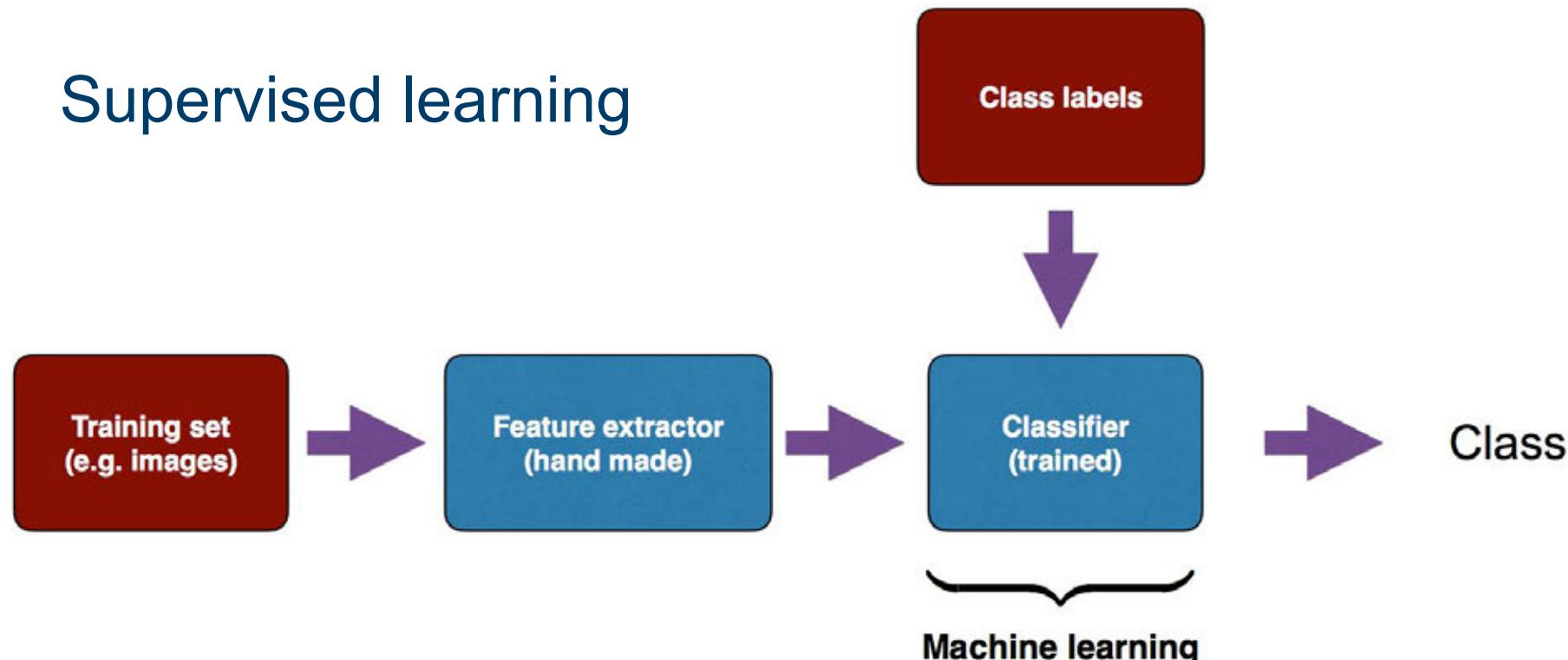
Image Analysis

Lecture 9.3 - Introduction to Machine Learning

Idar Dyrdal

Machine learning (Pattern recognition)

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



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
- ...



Credit: L. Lazebnik

Classification system

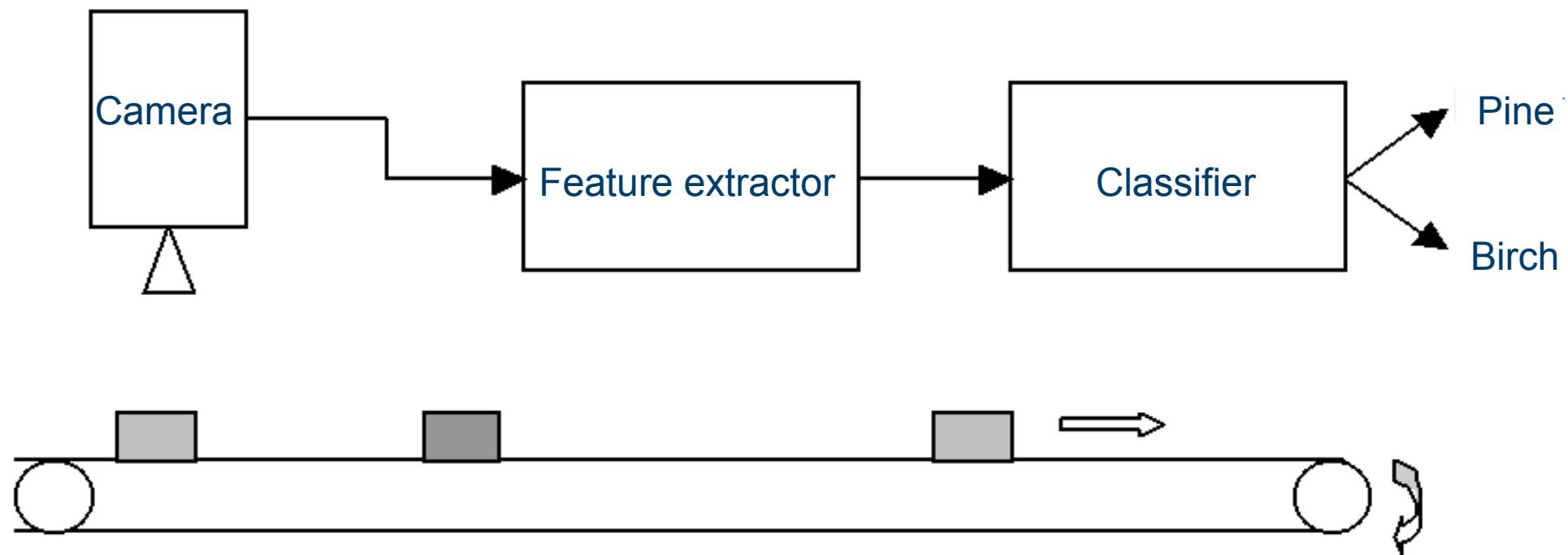


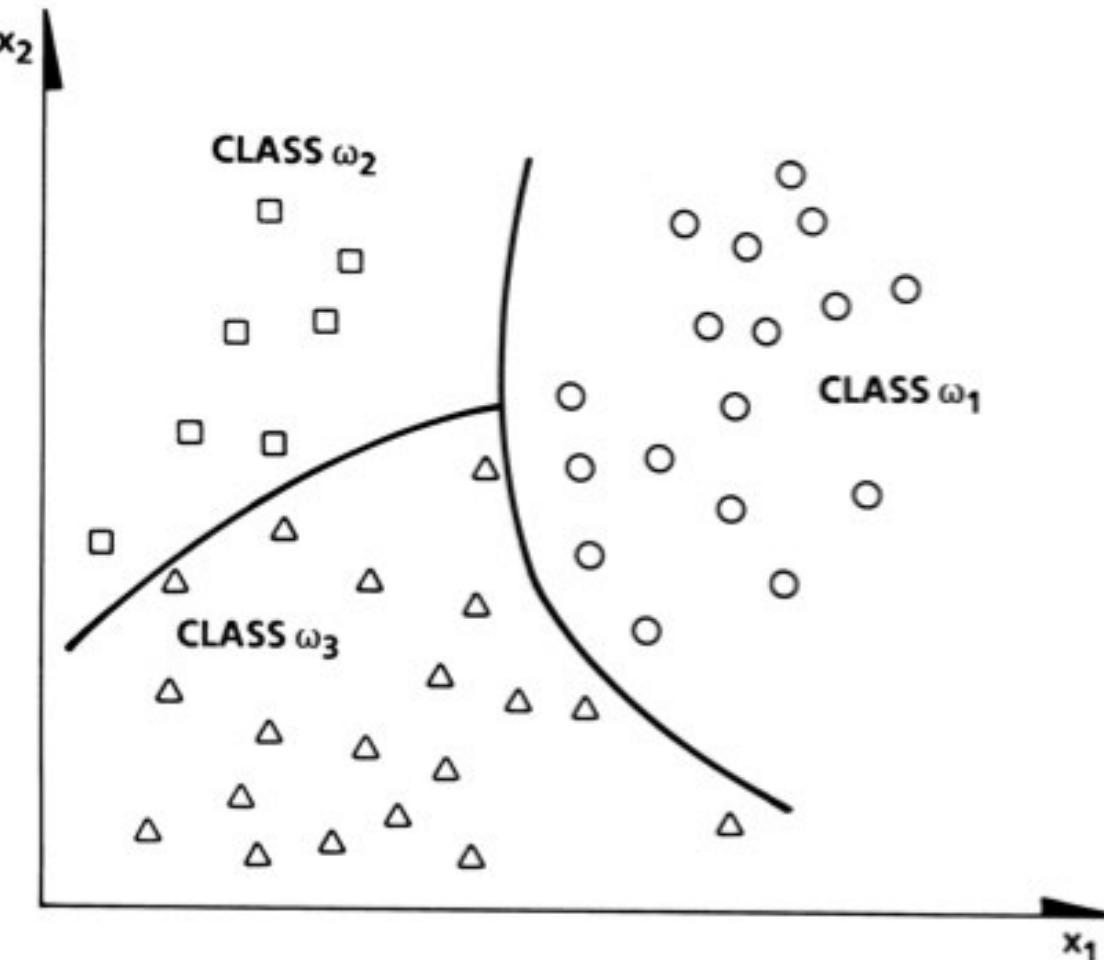
Image features for object recognition



$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix}$$

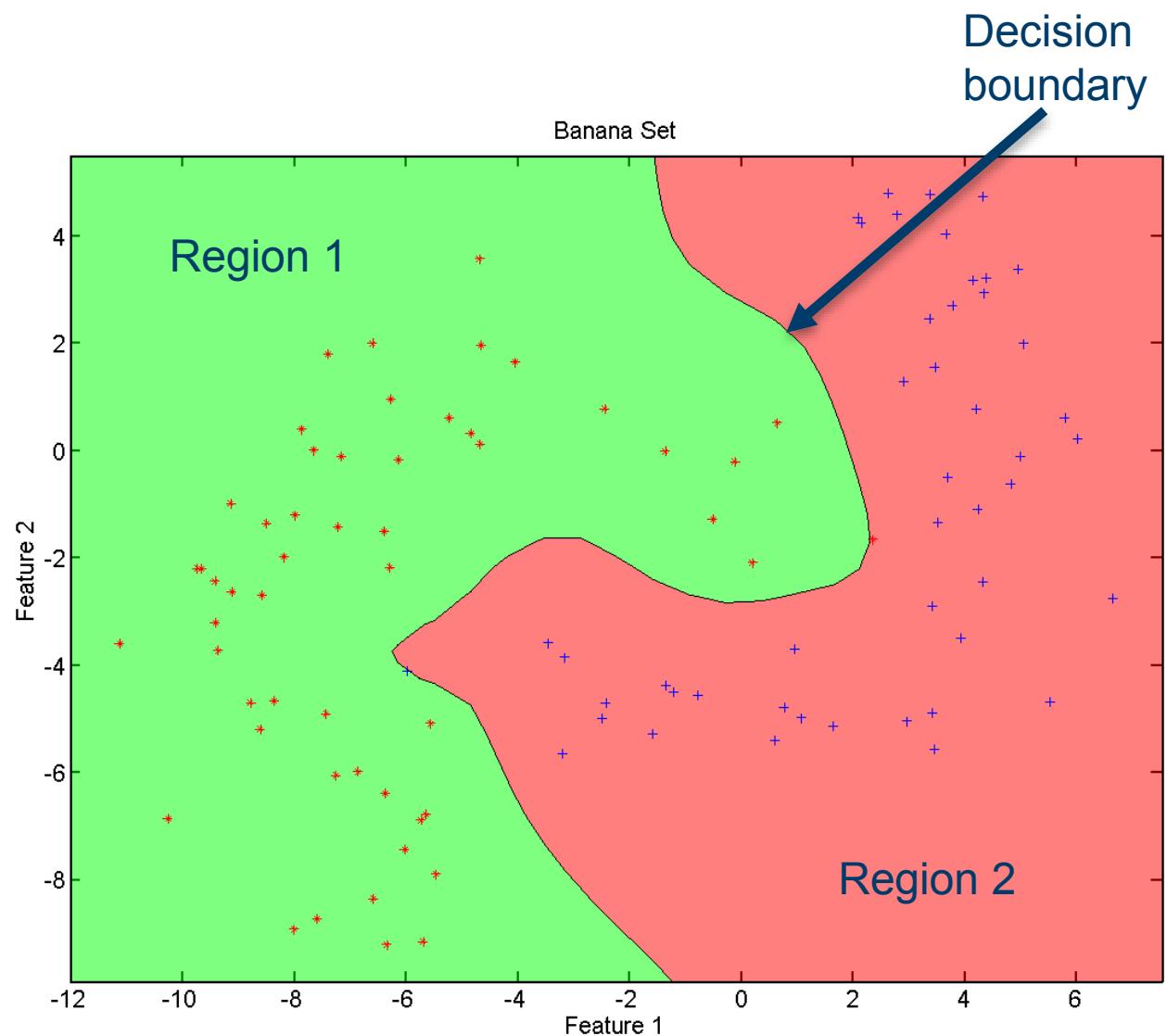
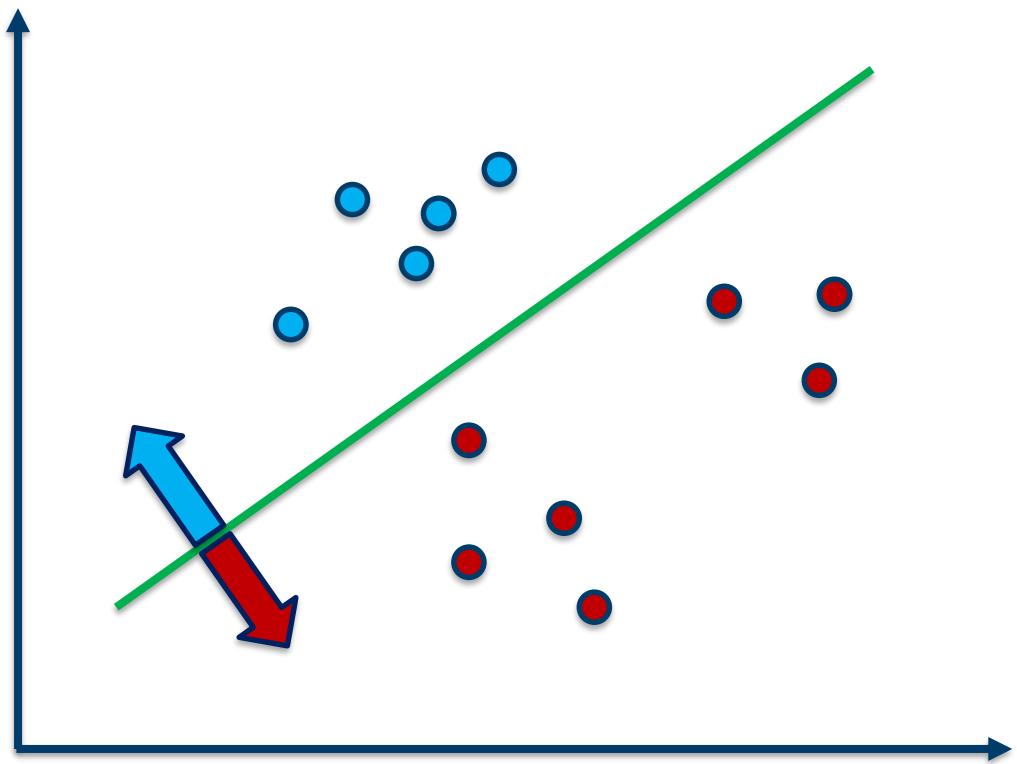
Feature vector and feature space

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_d \end{bmatrix}$$



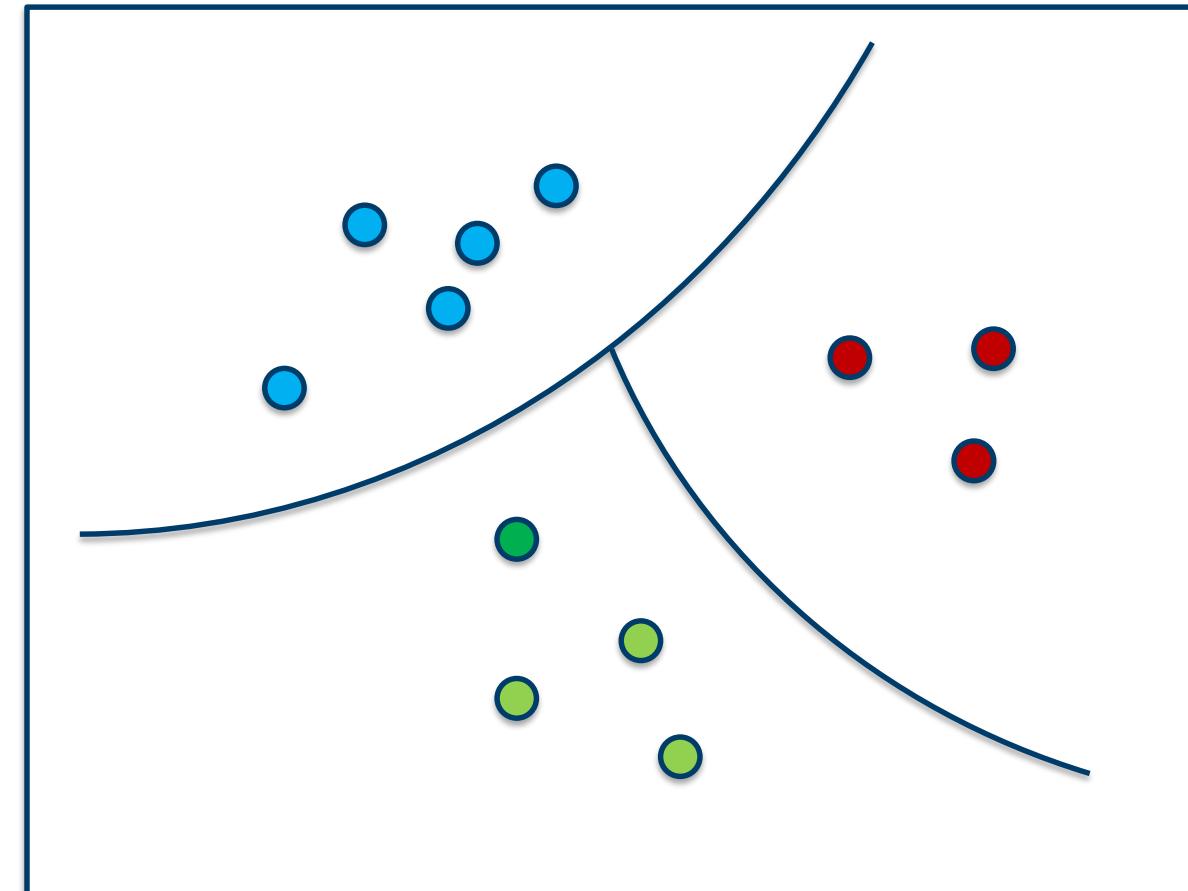
Training of classifiers

Learn a function to predict the class from the given features

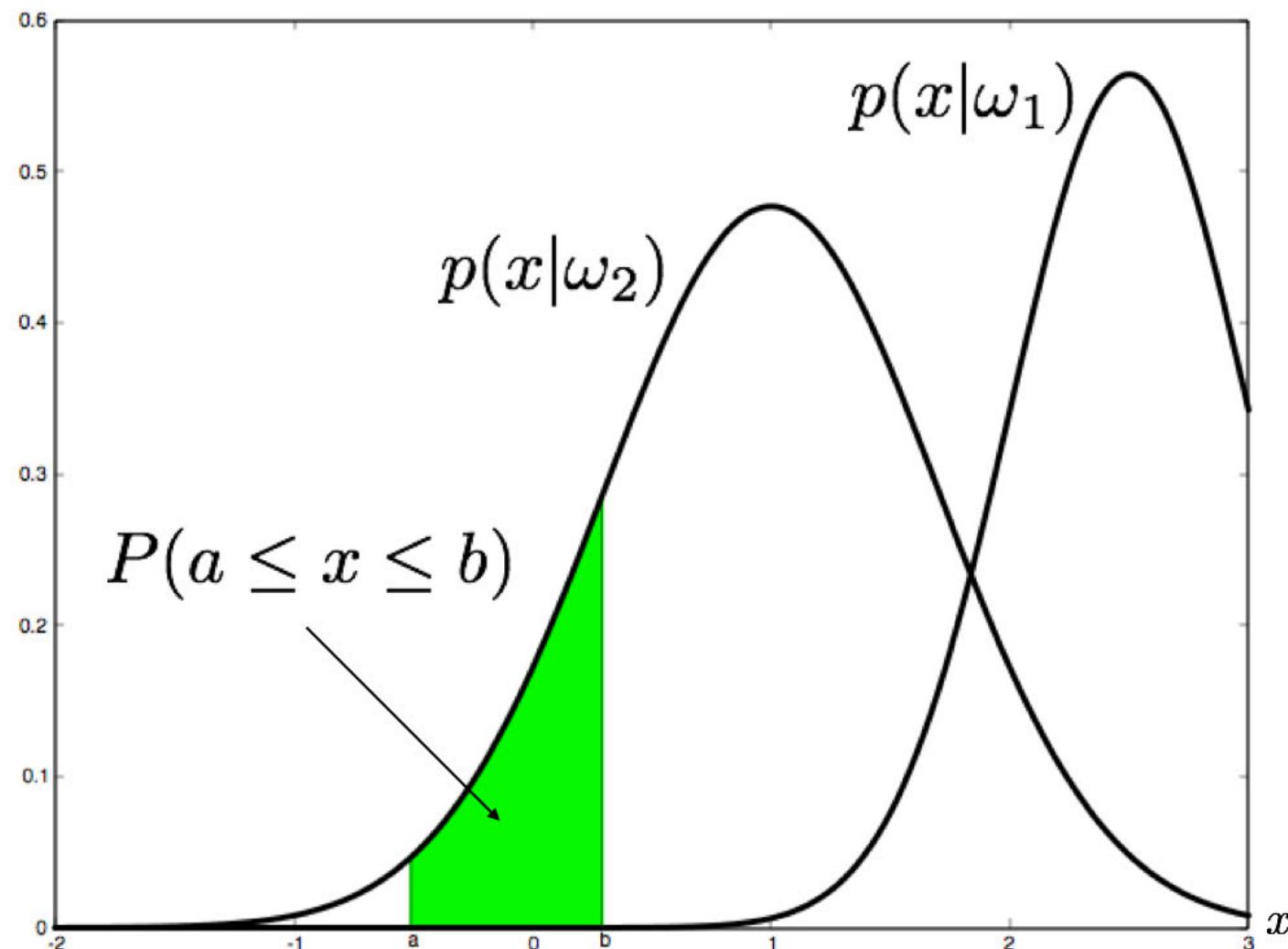


Classifiers and training methods

- Bayes classifier
- Nearest-neighbors and K-nearest-neighbors
- Parzen windows
- Linear and higher order discriminant functions
- Neural nets
- Support Vector Machines (SVM)
- Decision trees
- Random forest
- ...



Class conditional probability density functions



Bayesian decision theory

Overview

Class conditional densities:

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

Prior probabilities:

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

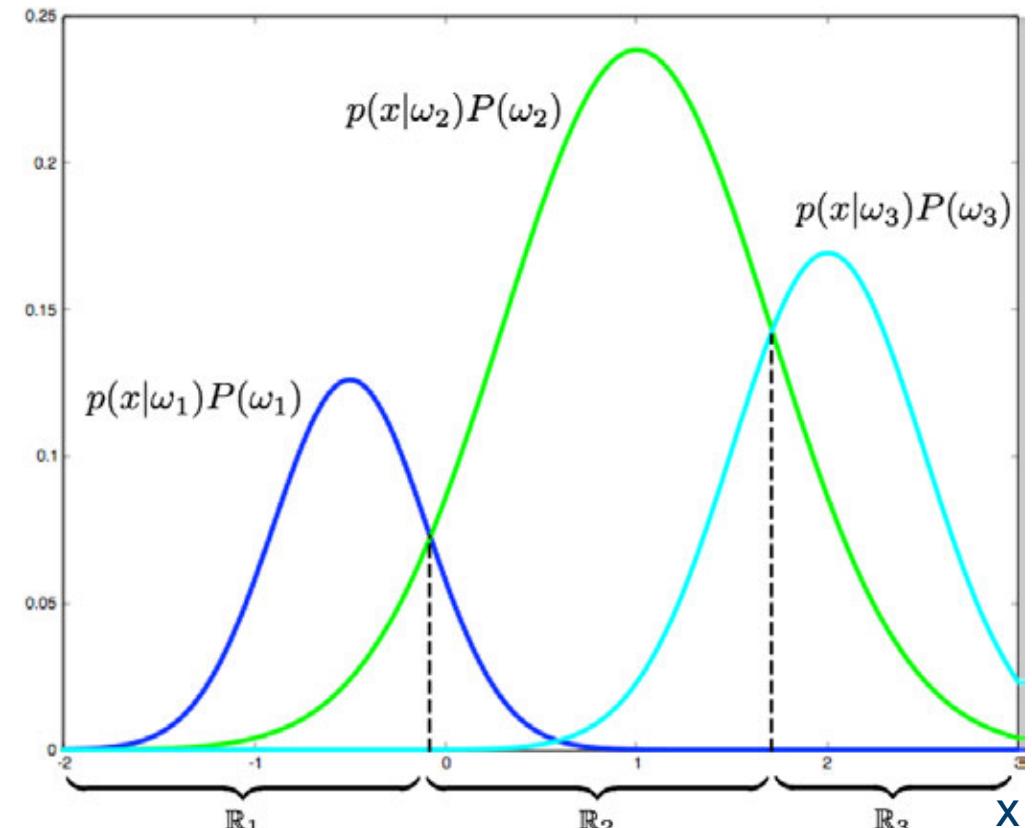
Posterior probabilities given by Bayes rule:

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

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

Minimum error rate classification:

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



Density estimation

Example – Gaussian distribution:

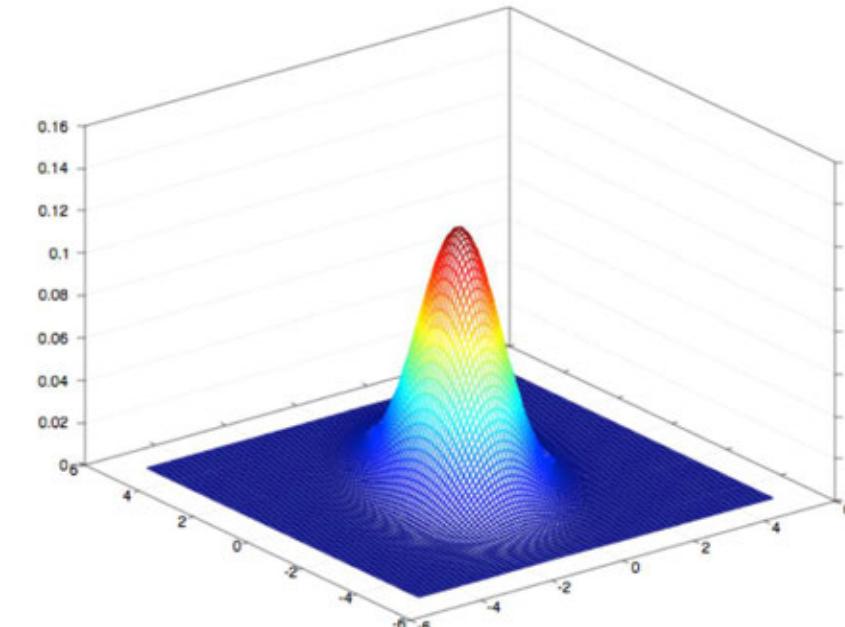
Parametric methods:

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

Non-parametric (distribution free)

methods:

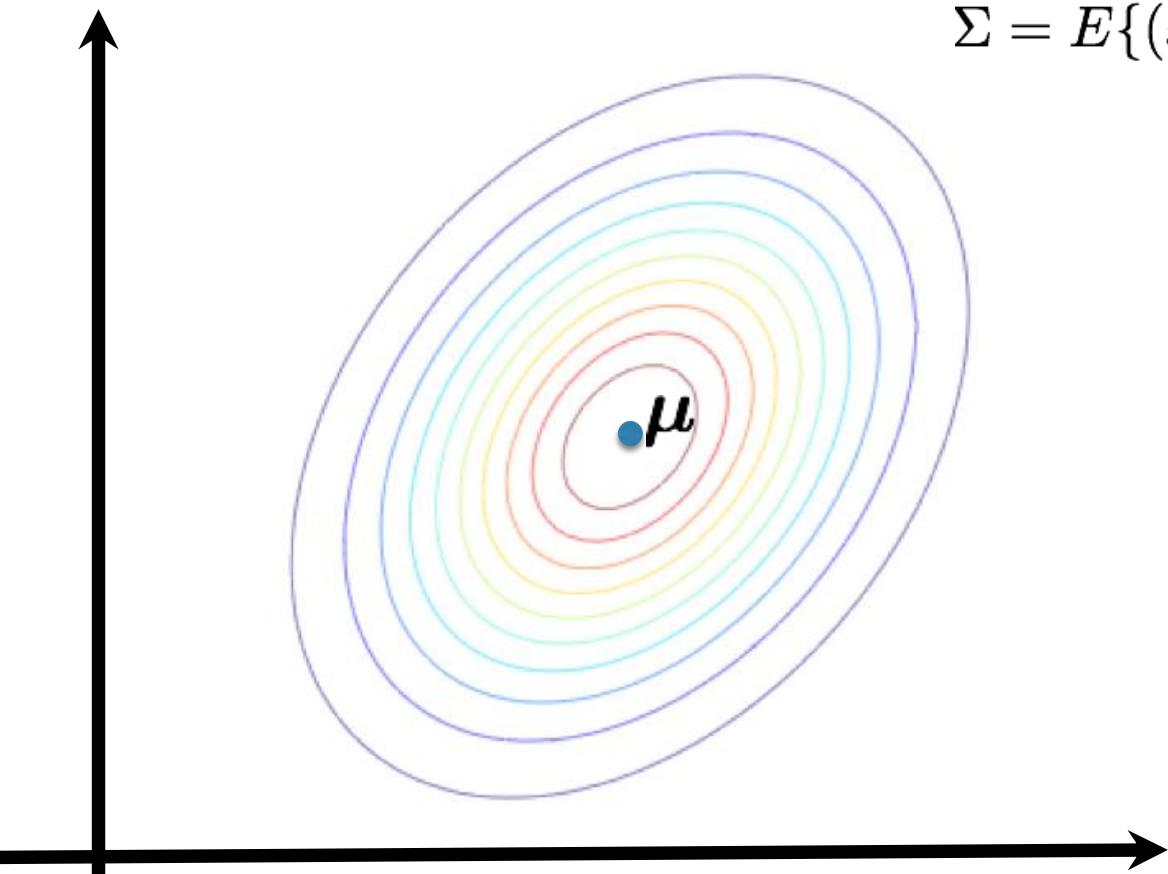
- Point estimation of the density using the training set directly
- Parzen windows
- Nearest neighbor estimation (leads directly to the nearest-neighbor and k-nearest-neighbor classifiers).



$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^t \Sigma_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right]$$

Parameters: $\boldsymbol{\mu}_i$ and Σ_i

Parameter estimation



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

Parameter estimates:

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

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

Discriminant functions

Estimate of the density in a given point:

$$\hat{p}(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\hat{\Sigma}_i|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \hat{\mu}_i)^t \hat{\Sigma}_i^{-1} (\mathbf{x} - \hat{\mu}_i) \right]$$

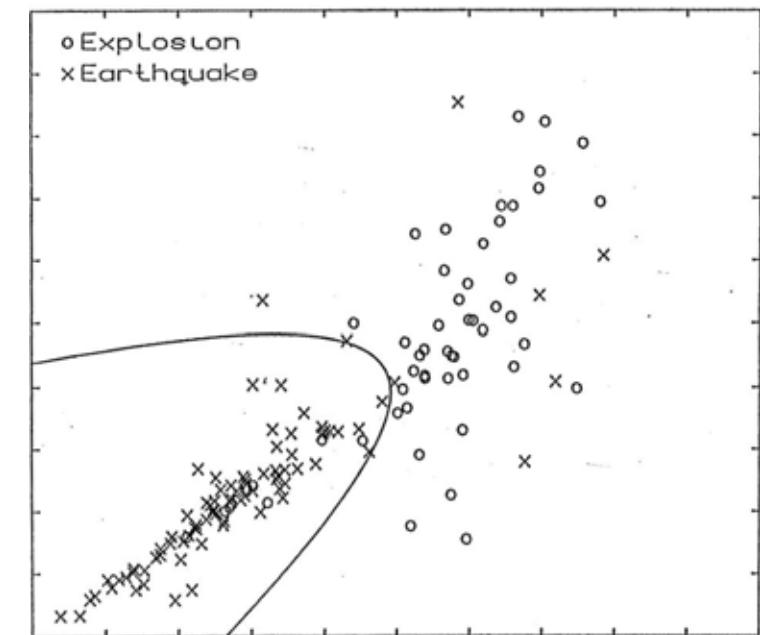
From Bayes rule:

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

Examples of discriminant functions:

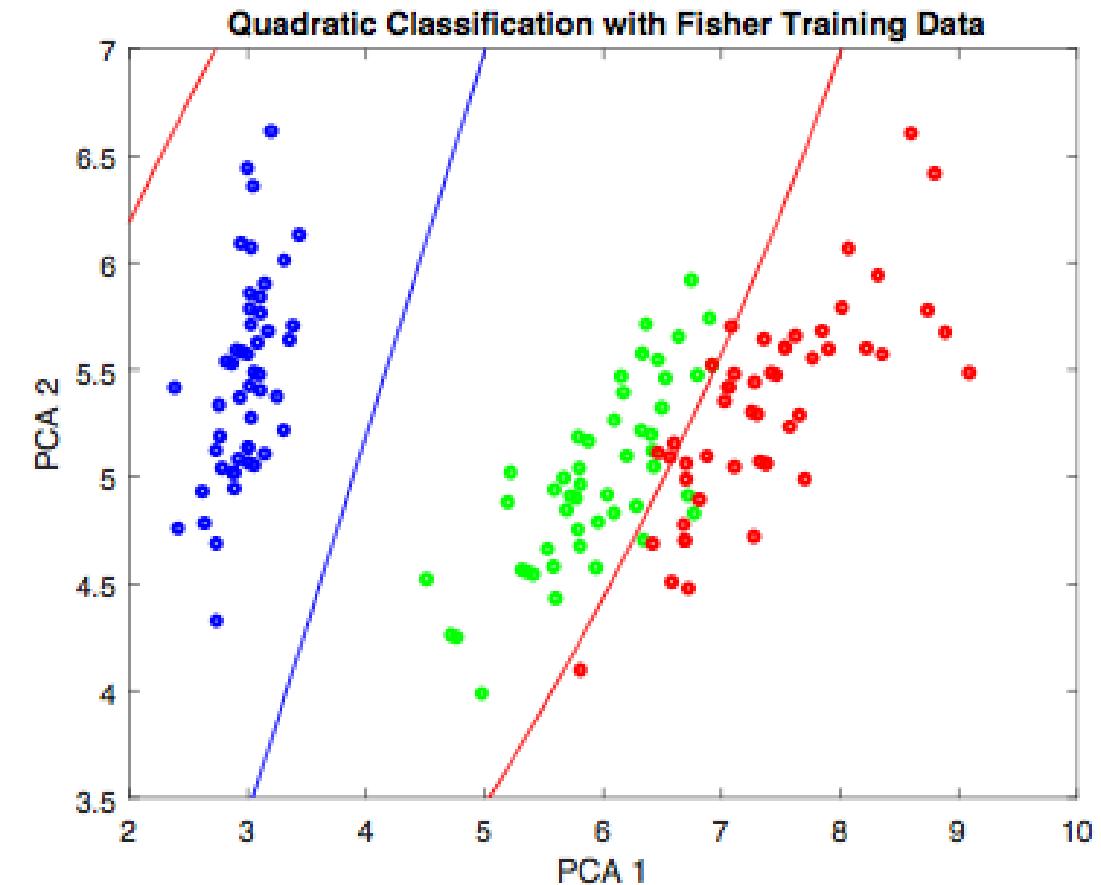
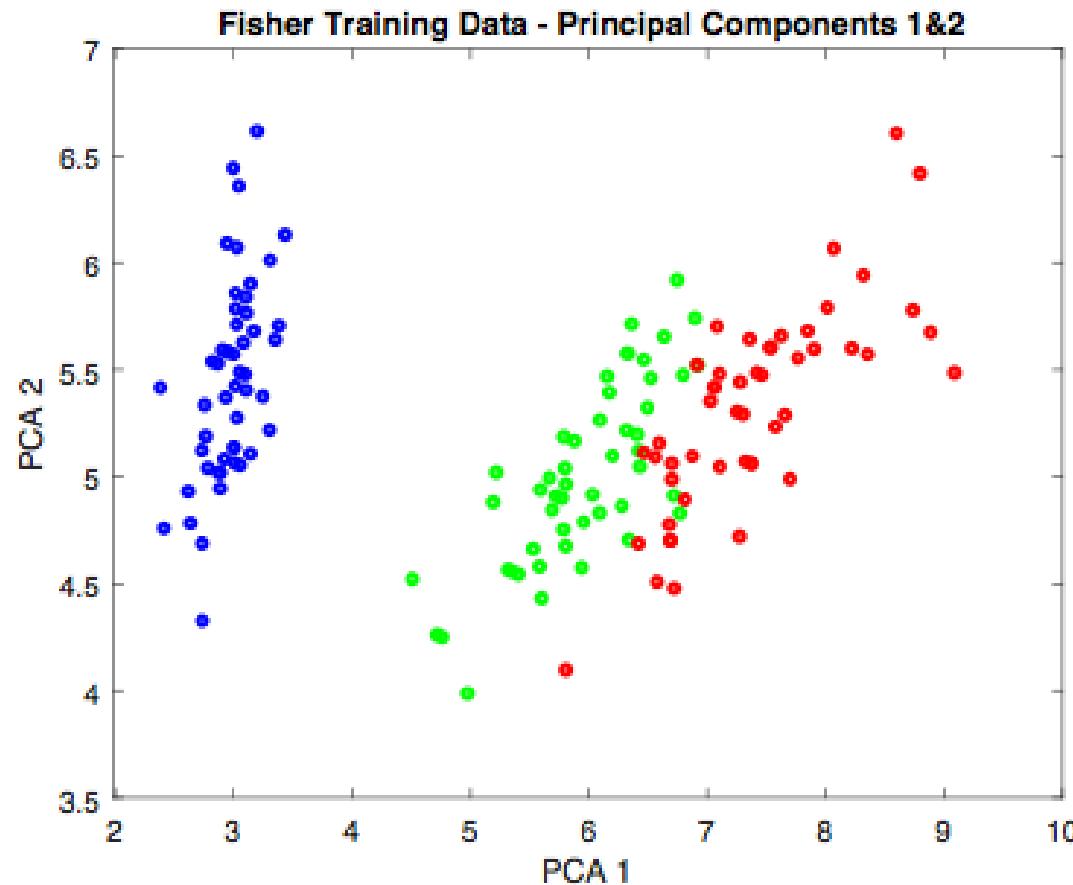
$$g_i(\mathbf{x}) = \ln \hat{P}(\omega_i | \mathbf{x}) \quad \text{or} \quad g_i(\mathbf{x}) = \ln \hat{p}(\mathbf{x} | \omega_i) + \ln P(\omega_i)$$

Decision rule:

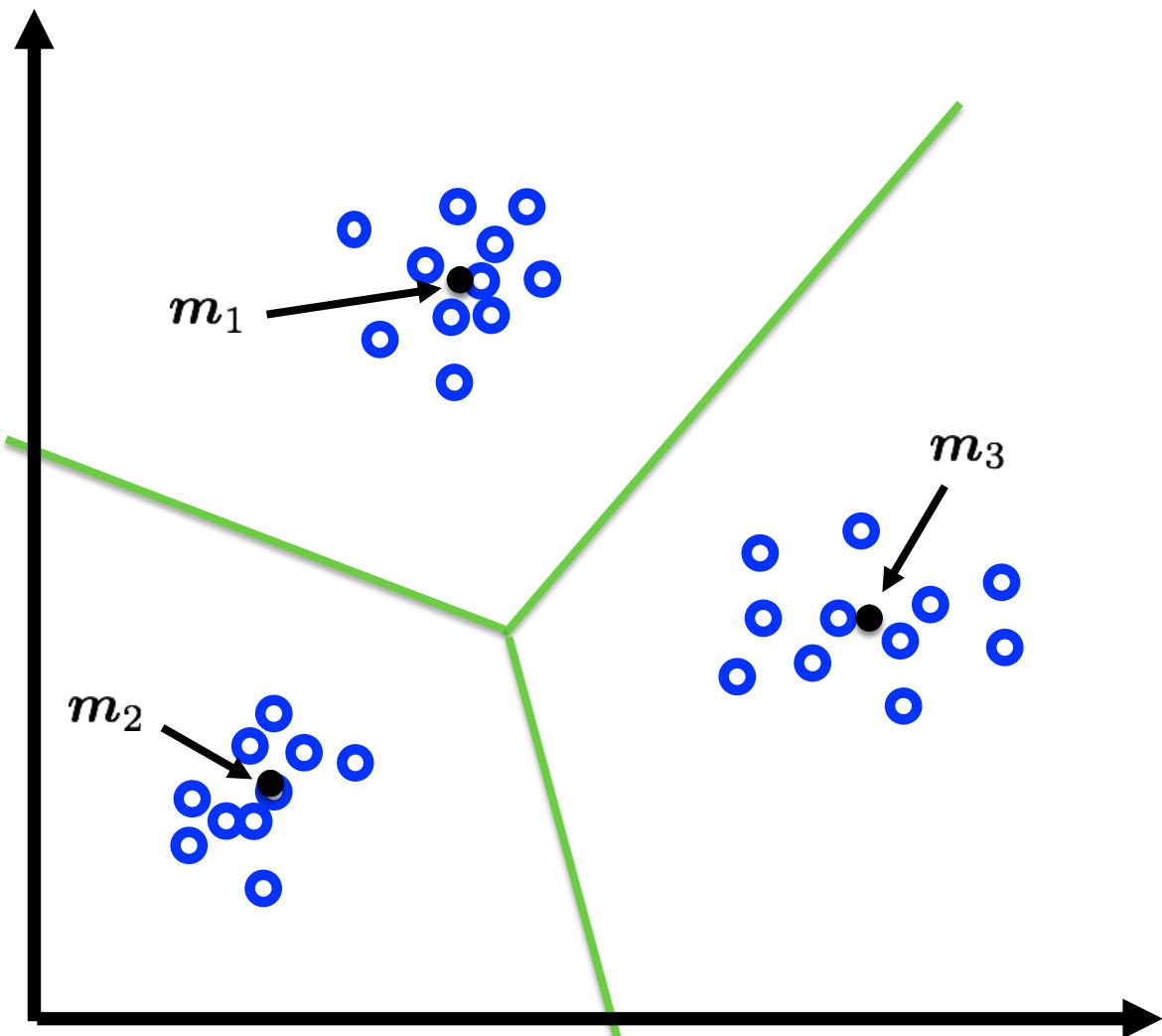


Choose the class with maximum discriminant function value.

Quadratic classifier - example



Linear classifier



Example:

Uncorrelated features and
common covariance matrices

Linear decision boundaries

Summary

Machine learning:

- Pattern classification
- Training of classifiers (supervised learning)
- Parametric and non-parametric methods
- Discriminant functions
- Quadratic and linear classifiers.

More information:

- Szeliski 14.1
- R. O. Duda, P. E. Hart, D. G. Stork (2001).
Pattern classification (2nd ed.). Wiley, New York. ISBN 0-471-05669-3.

