Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier
Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

This is a printable version of the slides of the lecture

Pattern Recognition (PR)
Winter term 2020/21
Friedrich-Alexander University of Erlangen-Nuremberg.

Erlangen, January 8, 2021
Prof. Dr.-Ing. Andreas Maier

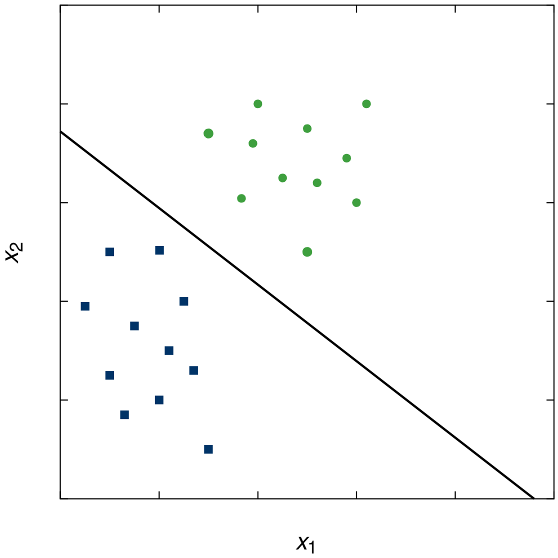
Support Vector Machines I

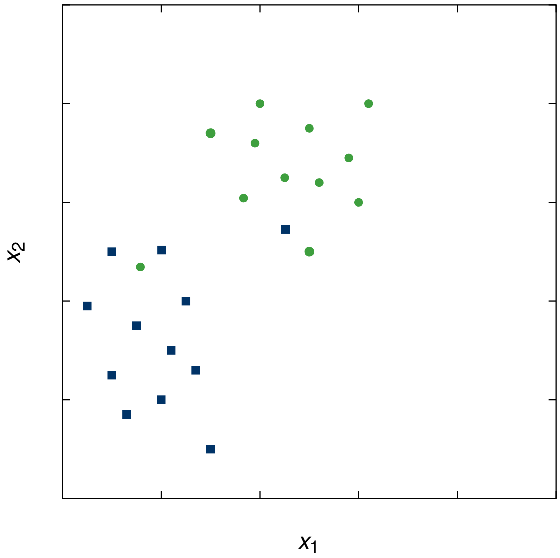
Motivation

- Assume two linearly separable classes.
- Computation of linear decision boundary that allows the separation of training data and that generalizes well.
- Vapnik 1996: Optimal separating hyperplane separates two classes and maximizes the distance to the closest point from either class. This results in
  - unique solution for hyperplanes, and
  - (in most cases) better generalization.

Motivation (cont.)

Linearly separable and non-separable classes

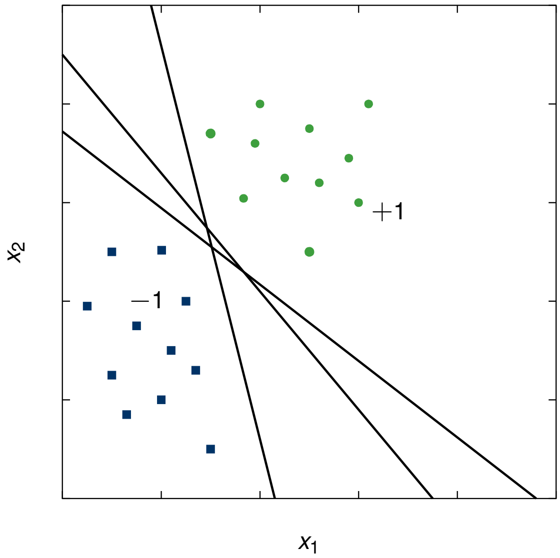




Pattern Recognition Lab
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Motivation (cont.)

Many, many, many solutions . . .

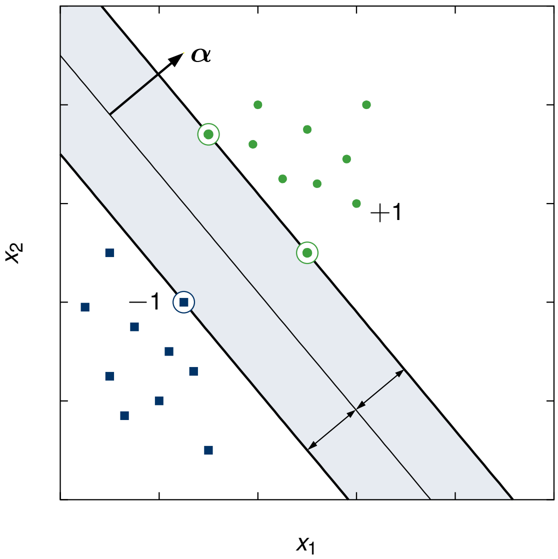


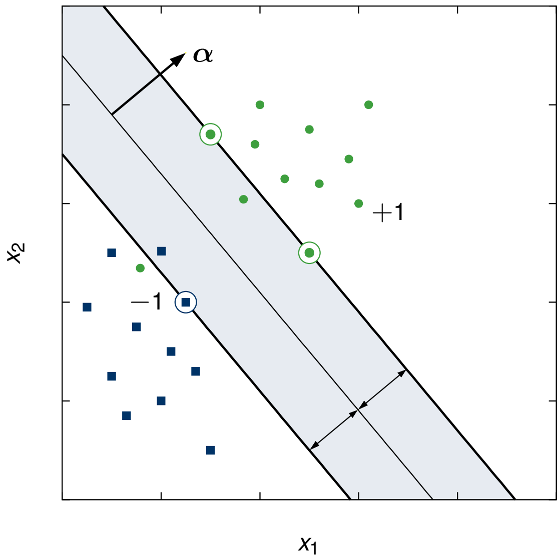
Idea: Average the perceptron solutions.

Motivation (cont.)

We distinguish between:

1. Hard margin problem
2. Soft margin problem





Remarks on Linear Algebra

Assume we have an affine function that defines the decision boundary:

$$f(\boldsymbol{x}) = \boldsymbol{\alpha}^T \boldsymbol{x} + \alpha_0$$

- For any point $\boldsymbol{x}$ on the hyperplane, we have $f(\boldsymbol{x}) = 0$.
- A necessary condition for two points on the hyperplane is:

$$f(\boldsymbol{x}_1) = f(\boldsymbol{x}_2) \quad \text{and thus} \quad \boldsymbol{\alpha}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 0.$$

- The normal vector $\boldsymbol{n}$ of the hyperplane is $\boldsymbol{n} = \boldsymbol{\alpha}/\|\boldsymbol{\alpha}\|_2$.

Remarks on Linear Algebra (cont.)

- The signed distance d of a point $\boldsymbol{x}$ to the hyperplane is:

$$d = \frac{\|\boldsymbol{\alpha}\|_2}{\boldsymbol{\alpha}^T\boldsymbol{\alpha}} \cdot f(\boldsymbol{x}) = \frac{1}{\|\boldsymbol{\alpha}\|_2} \cdot f(\boldsymbol{x}) = \frac{1}{\|\nabla f(\boldsymbol{x})\|_2} \cdot f(\boldsymbol{x})$$

- Assume points $\boldsymbol{x}_1, \boldsymbol{x}_2$ on either side of the margin satisfy $f(\boldsymbol{x}_1) = +1$ and $f(\boldsymbol{x}_2) = -1$.

  Thus we have:

$$\boldsymbol{\alpha}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2) = 2 \quad \text{and} \quad \frac{\boldsymbol{\alpha}^T}{\|\boldsymbol{\alpha}\|_2}(\boldsymbol{x}_1 - \boldsymbol{x}_2) = \frac{2}{\|\boldsymbol{\alpha}\|_2}$$

Constrained Optimization Problem

Constraints:

- Separation of classes has to be done with margin:

$$\begin{array}{lcll} \boldsymbol{\alpha}^T\boldsymbol{x}_i + \alpha_0 & \leq & -1, & \text{if} \quad y_i = -1 \\ \boldsymbol{\alpha}^T\boldsymbol{x}_i + \alpha_0 & \geq & +1, & \text{if} \quad y_i = +1 \end{array}$$

- This is equivalent to:

$$y_i \cdot (\boldsymbol{\alpha}^T\boldsymbol{x}_i + \alpha_0) \geq 1$$

Constrained Optimization Problem (cont.)

The maximization of the margin corresponds to the following optimization problem with linear constraints:

$$\begin{aligned} &\text{maximize} && \frac{1}{\|\boldsymbol{\alpha}\|_2} \\ &\text{subject to} && \forall i: \; y_i \cdot (\boldsymbol{\alpha}^T \boldsymbol{x}_i + \alpha_0) \geq 1 \end{aligned}$$

Note:

- Linear constraints ensure that all feature vectors have maximum distance to decision boundary.
- Basically we compute the distance of the convex hulls of feature sets.
- We need constrained optimization methods to solve the problem.

Constrained Optimization Problem (cont.)

The optimization problem is equivalent to

$$\begin{aligned} &\text{minimize} && \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2 \\ &\text{subject to} && \forall i: \; y_i \cdot (\boldsymbol{\alpha}^T \boldsymbol{x}_i + \alpha_0) - 1 \geq 0 \end{aligned}$$

Constrained Optimization Problem (cont.)

Remarks on the optimization problem:

- Convex optimization problem
- Efficient algorithms for solving the convex optimization problem (interior point method)
- Standard libraries can be used for minimization
- Solution is unique

Non-linearly Separable Classes

If classes are not linearly separable, we have to introduce *slack variables*.

Convex optimization problem:

minimize $\frac{1}{2}\|\alpha\|_2^2 + \mu \sum_i \xi_i$

subject to $\forall i: \ -(y_i \cdot (\alpha^T \boldsymbol{x}_i + \alpha_0) - 1 + \xi_i) \leq 0$,

$\forall i: \ -\xi_i \leq 0$

Pattern Recognition Lab

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG FACULTY OF ENGINEERING

Lessons Learned

- Support vector machine
- Hard and soft margin problem
- Convex optimization

Pattern Recognition Lab

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG TECHNISCHE FAKULTÄT

Next Time in

Pattern Recognition

Further Readings

- Bernhard Schölkopf, Alexander J. Smola:
  Learning with Kernels,
  The MIT Press, Cambridge, 2003.
- Vladimir N. Vapnik:
  The Nature of Statistical Learning Theory,
  Information Science and Statistics, Springer, Heidelberg, 2000.
- S. Boyd, L. Vandenberghe:
  Convex Optimization,
  Cambridge University Press, 2004.
  ☞ `http://www.stanford.edu/~boyd/cvxbook/`

Comprehensive Questions

- What is the concept of a SVM?
- What is the difference between a hard and soft margin SVM?
- What is the convex optimization problem of the hard margin SVM?
- What is the convex optimization problem of the soft margin SVM?