Project in Data Intensive Systems

4DV652

Lab Lecture 6

Welf Löwe

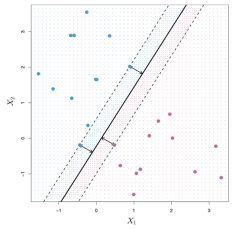
1

Agenda

- Support Vector Machines (SVM)
- Parameter optimization
- Lab 6 task descriptions

2

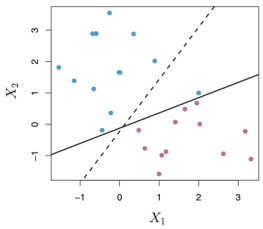
Maximal margin classifier



- Find a separating hyperplanes separating the training data points
  - Hyperplane is defined by a vector perpendicular to the plane
- Among all hyperplane, chose the one that maximizes the distance to the closest training data points

3

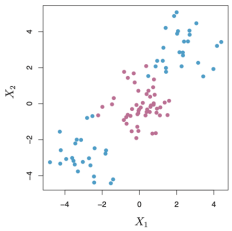
Support vector classifier (soft margin)



- Variance-bias tradeoff: we might choose the classifier that does a classification error (dashed line)
- Among all hyperplane, chose the one that
  - maximizes the distance to the closest training data points
  - minimizes the classification error
- Choose a a tuning parameter C to control the tradeoff

4

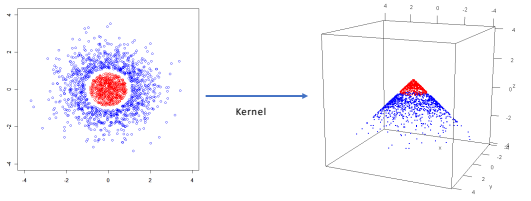
Support vector machines (SVM)



- Sometimes, no hyperplane is a good separator
- Then, apply a kernel function that transforms the data
  - Choose between different kernel functions
  - Choose the parameter values of the selected kernel functions
- Find a soft margin classifier in the transformed data space
  - Corresponds to a non-linear classifier in the original space

5

A so-called kernel transforms the datapoints

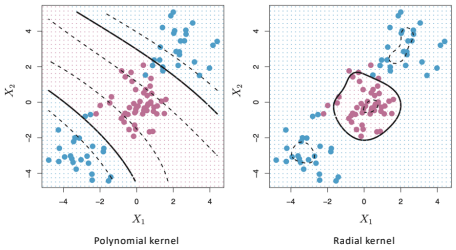


Red and blue points are not linearly separable

Each point x gets an extra coordinate $z=1-\|x\|)$, then the problem becomes linearly separable: the red and blue points can be separated by a 2-dimensional plane through $z=0$

6

Different kernels (functions) lead to nonlinear classification boundaries, here two example



7

More than two classes

As discussed before in the context of logistic regression:

- One-versus-all: learn r models
  - Each predicting the probability of class c_i against the probability of not c_i
  - Choose the class with the highest probability
- One-versus-one: learn all $r(r\text{-}1)/2$ models comparing pairs of predictors
  - Each predicting the probability of class c_i against the probability of another class c_j
  - Choose the class that wins a "KO tournament" or the most pairwise comparisons

8

Agenda

- Support Vector Machines (SVM)
- Parameter optimization
- Lab 6 task descriptions

9

Select model parameters

- Parameter selection is a general problem in classifiers
- In SVM, there a many parameters:
  - Tuning parameter C
  - Kernel function: polynomial, radial, …
  - Each kernel function has its own parameters
- Automated optimization instead of manual tuning

10

Grid optimization

- Nested loops, one for each parameter, iterate over parameter values
  - Quantitative parameter: choose all values (e.g., all kernels)
  - Qualitative parameters: define a grid of values (range and stride)
- In the loop body, assess the current model
  - Yet another loop for cross-validation
- The right model is yet another quantitative parameter with parameter values: logistic, k-means, Bayes, SVM, …
  - Selected in an outermost loop as it governs the other parameters
- Mind the optimization time
  - restrict the number of grid points if necessary
  - adaptive grid optimization (more grid points closer to optimum)

11

SVM grid optimization

- Normalize the predictors so that they are in the same value range
- Which kernel should we use and with which parameters? E.g.
  - if radial kernel, which γ?
  - if polynomial kernel, which a, b, c?
- What is the best parameter C trading max margins off against min classification error?
- The best kernel and the best combination of C and kernel parameters, e.g., γ, are selected by a grid optimization with exponentially growing values, e.g., $C \in \{2^{-5}, 2^{-3}, \ldots, 2^{13}, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-13}, \ldots, 2^{1}, 2^{3}\}$
- Each combination of parameter choices is checked using cross validation,
- Kernel and the parameters with best cross-validation accuracy are picked.

12

Agenda

- Support Vector Machines (SVM)
- Parameter optimization
- Lab 6 task descriptions

13

Lab assignment 6: SVM and Grid Optimization

- ML
  - Consider SVM classification as an alternative to the classification endpoint
  - Implement grid optimization for selecting the champion approach for this endpoint
  - Hint: https://github.com/WelfLowe/ML4developers/ notebook 5
  - Challenge the current champion classification with the SVM approach
- Software development
  - Consider the most accurate of each implemented classification model: logistic, *k*-means, Bayes, SVM, …
  - Assess the classification response time of these different classifiers
  - If applicable, deploy the SVM classifier as the new implementation of the classification endpoint
- Reporting:
  - In a sixth notebook, document the iteration(s) over the ML process steps
  - Report classification response of the alternative classifiers
  - Prepare a short talk/presentation of the project (20-30 min), 10-15 min for your team organization, another 10-15 min presenting ML/tech findings.
- Deadline: 2023-03-15

14