

IN 5520 Weekly exercises on Support Vector Machines.

Exercise 1.

$y_i(w^T x_i + w_0) \geq 1, \quad i = 1, 2, \dots, N$ Show that the criterion

corresponds to correct classification for all N samples in a binary classification problem with classes -1 and 1.

Exercise 2.

Given a binary data set:

$$\text{Class - 1: } \begin{bmatrix} 1 & 9 \\ 5 & 5 \\ 1 & 1 \end{bmatrix} \quad \text{Class + 1: } \begin{bmatrix} 8 & 5 \\ 13 & 1 \\ 13 & 9 \end{bmatrix}$$

Plot the points in a plot. Sketch the support vectors and the decision boundary for a linear SVM classifier with maximum margin for this data set.

Exercise 3.

Given the binary classification problem:

$$\text{Class -1: } \begin{bmatrix} 2 & 2 \\ 3 & 3 \\ 4 & 4 \\ 5 & 5 \\ 4 & 6 \\ 3 & 7 \\ 4 & 8 \\ 5 & 9 \\ 6 & 10 \end{bmatrix} \quad \text{Class + 1: } \begin{bmatrix} 6 & 2 \\ 7 & 3 \\ 8 & 4 \\ 9 & 5 \\ 8 & 6 \\ 7 & 7 \\ 7 & 8 \\ 7 & 9 \\ 8 & 10 \end{bmatrix}$$

- Sketch the point in a scatterplot.
- In the plot, sketch the mean values and the decision boundary you would get with a Gaussian classifier with $\Sigma = \sigma I$.
- What is the error rate of the Gaussian classifier on the training data set?

- d) Sketch on the plot the decision boundary you would get using a SVM with linear kernel and a high cost of misclassifying training data. Indicate the support vectors and the decision boundary on the plot.
- e) What is the error rate of the linear SVM on the training data set?

Exercise 4.

Download the two datasets `mynormaldistdataset.mat` and `mybananadataset.mat` from `undervisningsmateriale/images`.

HINT: If you want a starting code, use `svm_hints.m` and play with the parameters.

You can use a library for SVM e.g. `fitcsvm` to train and `svmclassify` in Matlab

Familiarize you with the data sets by studying scatterplots.

Load `mynormaldistdataset.mat`. Stick with the linear SVM, but change the C-parameter ('BoxConstraint').

Rerun the experiments a couple of times, and visualize the data using 'ShowPlot'. How does the support vectors and the boundary change with the parameter?

Try to remove some of the non-support-vectors and rerun – does the solution change?

Load `mybananadataset.mat`. Try various values values of the C-parameter with a linear SVM. Can the linear SVM classifier make a good separation of the feature space?

Change kernel to a RBF (radial basis function), and rerun. Try changing the sigma-parameter ('rbf_sigma' in `svmtrain`). Make sure you know why we now get a non-linear decision boundaries.

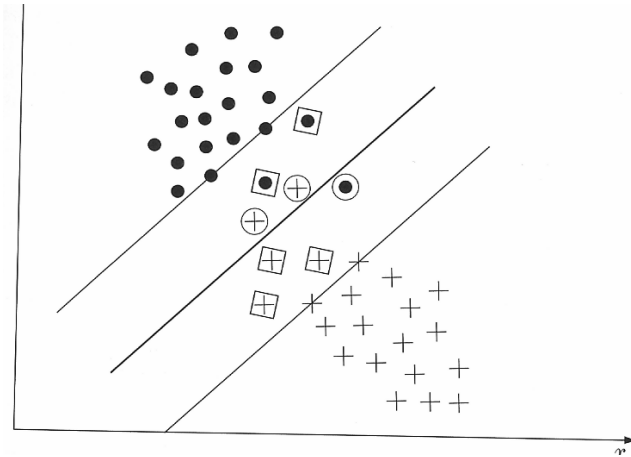
Implement a grid search of the C- and sigma-parameters based on 10-fold crossvalidation of the training data (the A-dataset). Find the best values of C and sigma, retrain on the entire A-data set, and then test on the B-data set. Does the average 10-fold crossvalidation estimate of the overall classification error match the result we get when testing on the independent 'B'-dataset?

Use the parameter range listed in the lecture foils.

Exercise 5: Support vector machine classifiers (from 2017 exam)

- a) Consider a linear SVM with decision boundary $g(x) = w^T x + w_0$. In SVM classification, explain why it is useful to assign class labels -1 and 1 for a binary classification problem.

- b) The basic SVM optimization problem is to minimize $J = \frac{1}{2} \|w\|^2$
 What are the additional constraints for this optimization problem?
 Ideally, you should answer both by math and explain what this expression means.
- c) Explain how slack variables ξ_i are used to solve a non-separable case like the one below:



- d) Discuss how likely a Gaussian classifier and an SVM classifier are to overfit to the training data.
- e) Explain how an SVM can be used for a classification problem with M classes.
- f) Explain briefly how SVM parameters should be determined

Exercise 6 (from 2018 Exam) : Support vector machines

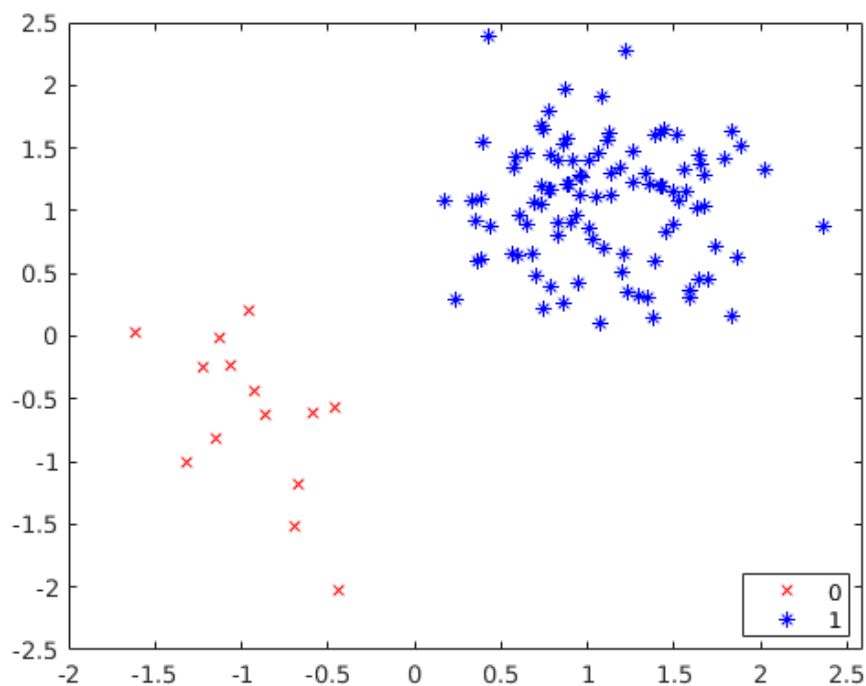
- a) The basic optimization problem for a support vector machine classifier is:

$$\text{minimize } J(w) = \frac{1}{2} \|w\|^2$$

$$\text{subject to } y_i(w^T x_i + w_0) \geq 1, \quad i = 1, 2, \dots, N$$

What is the total margin for this problem?

- b) Support vector machines are fundamentally different from Gaussian classifiers in terms of how the decision boundary is found – explain why.
- c) Support vector machine classifiers can also be explained based on convex hulls. Explain the relationship between the convex hull of two regions and the hyperplane with maximum margin.
- d) Given below is a scatter plot of a binary classification problem. Sketch the convex hulls on the figure and use this to find an approximate hyperplane.



e) In the general case the optimization problem is given as:

$$\max_{\lambda} \left(\sum_{i=1}^N \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j x_i^T x_j \right)$$

subject to $\sum_{i=1}^N \lambda_i y_i = 0$ and $0 \leq \lambda_i \leq C \quad \forall i$

Explain briefly which terms in the equation that can be computed using kernels in a high-dimensional space, and also explain what the kernels measure.