183.605
Machine Learning for Visual Computing
Assignment 2

Group 12:
Hanna Huber (0925230)
Lena Trautmann (1526567)
Elisabeth Wetzer (0726681)

January 30, 2017

1 Assignment 2

1.1 The dual optimization problem

Support Vector Machines, short SVM, maximize the distance between data points and the decision boundary, thereby cope with the common issue that new data points are easily misclassified if the decision boundary is located close to another data point. A linear decision boundary can be described by the equation $d(x) = w^T x + w_0 = 0$ with a vector $w \in \mathbb{R}^d$ and $w_0 \in \mathbb{R}$. The distance of a point $x \in \mathbb{R}^d$ to the decision boundary is given by $\frac{w^T x + w_0}{||w||}$. Furthermore let $d(x_i)t_i > 0$. The distance to the decision boundary is to be maximized which leads to:

$$\frac{d(x_i)t_i}{||w||} \geq \tau$$

for $i \in \{1, 2, ..., N\}$ and $\tau > 0$ the margin. Since the hyperplane equation $d(x) = 0$, w as well as w_0 can be multiplied by a positive scalar without manipulating the solution, τ can be assumed as $\frac{1}{||w||}$. The margin τ is maximized when the function $J = \frac{1}{2}||w||^2$ is minimized under the constraints of $(w^T x_i + w_0)t_i \geq 1$.
This can be achieved by using Lagrange multipliers, leading to the formulation:
Maximize:

$$L(\alpha) = -\frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j t_i t_j (x_i^T x_j) + \sum_{i=1}^{N} \alpha_i$$

under the constraints $\alpha_i > 0 \quad \forall i$ and $\sum_{i=1}^{N} \alpha_i t_i = 0$.

In other notation it can be expressed as minimizing:

$$-L(\alpha) = \frac{1}{2}\alpha^T H \alpha + f^T \alpha$$

Tasks:

- Generate a suitable training set of linearly separable data

The training set is generated in `generateTrainingData.m`. The function `generateTrainingData(N, xRange, yRange, linear)` takes the number of sample points, the domains (defined by a lower and an upper bound) from which the x and y coordinates are sampled and a flag which indicates if the resulting data should be linearly separable. A set of random 2D coordinates is created using the MATLAB function `rand`. Linear separability is achieved by labelling the points (x_i, y_i) according to the condition $x_i + y_i > \bar{x} + \bar{y}$, where $\bar{x}$ and $\bar{y}$ denote the domain centers.

- Plot the input vectors in $\mathbb{R}^2$ and visualize corresponding target values (e.g. by using color).

The resulting training data is displayed in Figure 1. Sample points with class label 1 are marked red, points with label -1 are marked green.

- Visualize the support vectors and plot the decision boundary.

Figure 2 shows the support vectors defined by `trainSVM` and the decision boundary.

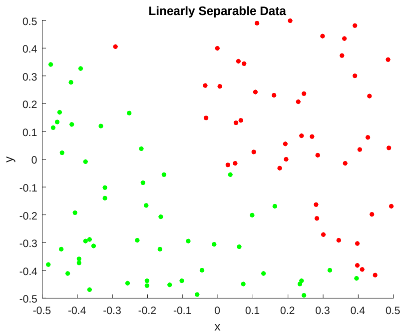


Figure 1: Linearly separable data with color-coded class labels (1=red, -1=green) for $N = 100$.

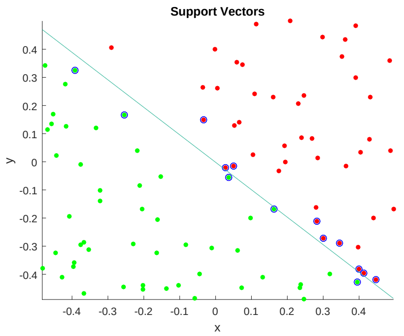
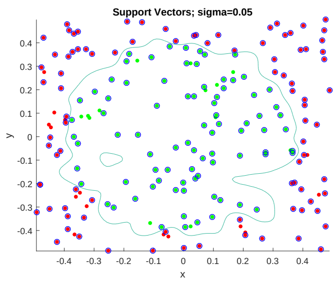


Figure 2: Training data with support vectors marked by blue circles and decision boundary plotted in blue.

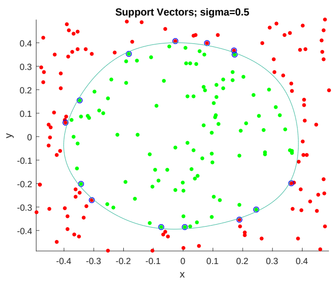
1.2 The kernel trick

Tasks:

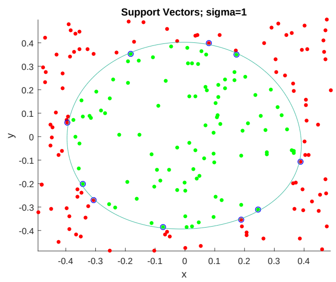
- Try different values for σ (the RBF parameter).



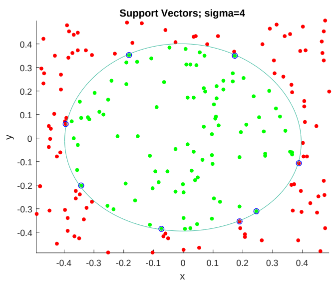
(a) $\sigma = 0.05$



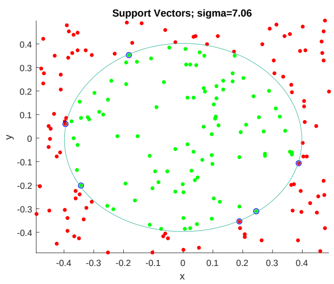
(b) $\sigma = 0.5$



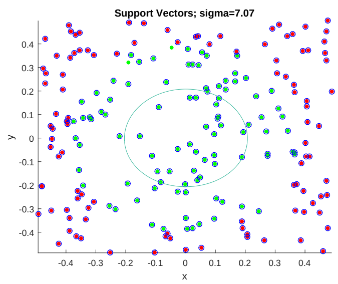
(c) $\sigma = 1$



(d) $\sigma = 4$



(e) $\sigma = 7.06$



(f) $\sigma = 7.07$

Figure 3: Training data with support vectors marked by blue circles and decision boundary plotted in blue, for different values of the RBF parameter σ.

The radial basis function (RBF) kernel is defined by $K(x, y) = exp(-\frac{\|x-y\|^2}{\sigma^2})$. Figure 3 shows the support vectors and the decision boundaries for different values of σ for the same non-linear separable dataset. We can see, that the decision boundary gets better for larger σ up to a value of 7.07. With $\sigma >= 7.07$ the values of α are all greater than 10^{-8}, hence we have all data points as support vectors. Furthermore, the least number of SVs was determined for $\sigma = 7.06$.

- Generate a non-linearly separable training set, plot the data, visualize the support vectors and plot the decision boundary.

Figure 3 shows that the decision boundaries for different values of σ, but the same non-linear seperable data set. It can be observed that a varying σ leads to a varying number of support vectors.

To train the non-linear seperable data using a RBF kernel using a slack variable C, we introduce an upper boundary for the Lagrange multipliers α_i. The upper boundary is given by $\frac{C}{N}$. The value of the slack variables needs to be chosen carefully to prevent overfitting in case of large slack variables but may be chosen large enough to provide accurate SVM calculations.