Pattern Recognition Winter Term 2019
## Institute of Neural Information Processing
PD Dr F. Schwenker

Assignment 5 (Submission until December 3, 2019)

Exercise 1 (4 points): Idea of Maximal Margin (SVM) [Pen and Paper]

With the help of support vector machines, we can find a separation line which maximize the margin between two classes. In this exercise, we use a two-class example and approach to such a line in a geometric way. Figure 1 shows the two-dimensional data we are using throughout this exercise. Each point $x \in \mathbb{R}^2$ belongs either to the positive ω_1 (blue, circles) or negative class ω_{-1} (orange, squares).

1. To find the separation line, the convex hull of a dataset is helpful. Start by drawing the convex hull for each class in Figure 1.

2. The line with the maximal margin is defined by the two points from the convex hulls of the two classes with minimal distance to each other. Find these two points on the convex hulls you have just drawn and label them with c_1 and c_{-1} (for the class ω_1 and ω_{-1}, respectively). Note that every point on a convex hull is a possible candidate and they do not necessarily need to correspond with the data points.

3. Connect the points c_1 and c_{-1} with a line.

4. The separation line goes through the centre of the line you have just drawn and is orthogonal to it, i.e. the two lines enclose a 90° angle. Draw the separation line s, the line l_1 which passes through the support vectors of ω_1 and the line l_{-1} which passes through the support vectors of ω_{-1}.

5. Add two arbitrary points from each class, i.e. $x_1, x_2 \in \omega_1$ and $x_3, x_4 \in \omega_{-1}$, to the feature space so that the separation line s found previously does not change.

6. Start fresh with Figure 2 and add a new data point x_5 which belongs to a class of your choice to the feature space so that the new margin between the two classes equals 1. As before, sketch the three lines $\tilde{l}_1$, $\tilde{l}_{-1}$ and $\tilde{s}$ based on the points $\tilde{c}_1$ and $\tilde{c}_{-1}$ on the convex hull.

7. In general, what is the difference between a line obtained via SVM and a line obtained via the perceptron learning algorithm?
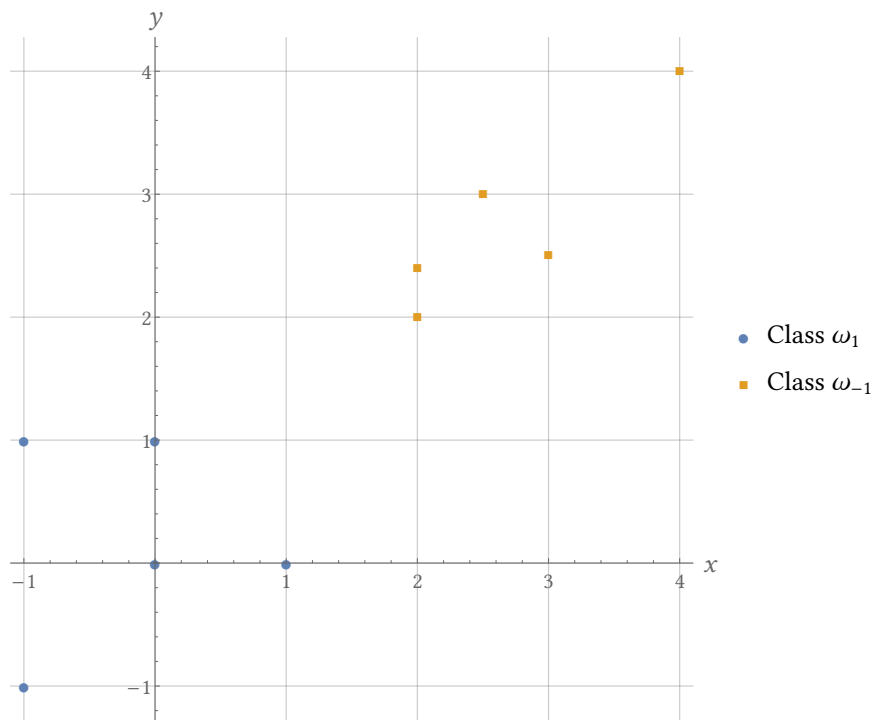
Figure 1: Example data of Exercise 1.
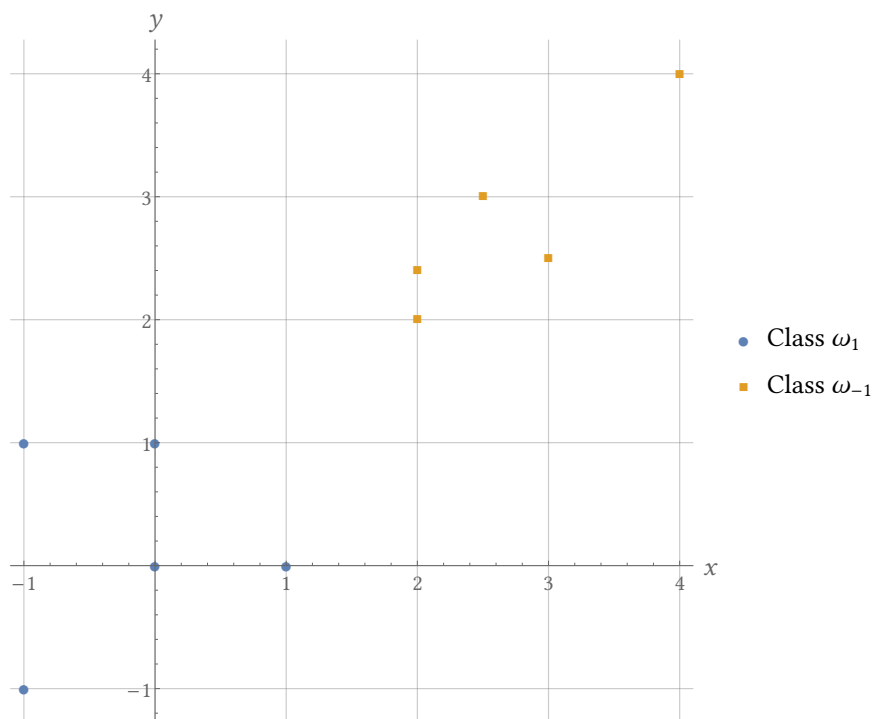
Figure 2: Example data of Exercise 1 (figure reprinted).

Exercise 2 (6 points): SVM [Pen and Paper]

From the previous exercise, we know which kind of separation line we can expect when using support vector machines and how to find it graphically. Determining the separation line analytically, however, is very hard; even for the simple problem of Exercise 1. The number of variables and conditions make this task impractical for a "Pen and Paper" exercise. But, to still get a basic idea of the SVM algorithm, we consider here an even simpler problem where we have only two points $x_i \in \mathbb{R}^2$ which belong to their own class each. The first point is from the set of positive samples $\omega_1 = \{x_i \mid T_i = 1\}$ and the second from the set of negative samples $\omega_{-1} = \{x_i \mid T_i = -1\}$. The two points are defined as

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in \omega_1 \quad \text{and} \quad x_2 = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \in \omega_{-1}.$$

Obviously, both points are also support vectors since they are the only representatives from their classes and hence must lie on the margin. Our goal is to find the parameters of a separation line with maximal margin, i.e. we search for w and w_0.

In order to find our parameters, we need to optimize the Lagrange function

$$L(w, w_0, \alpha_1, \alpha_2) = L(\alpha_1, \alpha_2) = \sum_{i=1}^{2} \alpha_i - \frac{1}{2} \sum_{i=1}^{2} \sum_{j=1}^{2} \alpha_i \alpha_j T_i T_j x_i^T x_j \tag{1}$$

subject to the constraint

$$\sum_{i=1}^{2} \alpha_i T_i = 0. \tag{2}$$

Technically, we also need the constraints $\alpha_i \geq 0$ (cf. page 100 in the script). But to keep things simple, we are assuming them satisfied here.

1. Set up the Lagrange function (Equation 1) as well as the constraint (Equation 2) of our problem set, i.e. plug in the values and simplify.

2. To optimize Equation 1 under the constraint of Equation 2, we need an additional Lagrange function with the Lagrange multiplier λ

$$\Lambda(\alpha_1, \alpha_2, \lambda) = L(\alpha_1, \alpha_2) + \lambda \left(\sum_{i=1}^{2} \alpha_i T_i \right).$$

Optimize this function and show that the optimal values are given as

$$\alpha_1^* = \frac{2}{5} \quad \text{and} \quad \alpha_2^* = \frac{2}{5}.$$

It is sufficient to calculate only the extrema values. You do not need to show that they belong to a minimum/maximum (and not a saddle point) of $L(\alpha_1, \alpha_2)$. You can use Figure 3 to convince yourself that this is indeed the case here.

3. Based on the previous results, calculate the line parameters w and w_0.

4. Draw the resulting separation line into Figure 4.

5. Suppose you want to classify the new point $x_3 = (2, 2)$. Assign this point the correct label

   a) graphically based on the line you have just drawn into Figure 4.

   b) analytically by using the decision function $f(x)$ (script page 102).

6. Some general questions to conclude this exercise (no calculations required):

   a) How would Equation 1 change if we had more than three points?

   b) How can we infer from the coefficient α_i^* whether the corresponding point x_i contributes to the separation line (i.e. whether x_i is a support vector)?
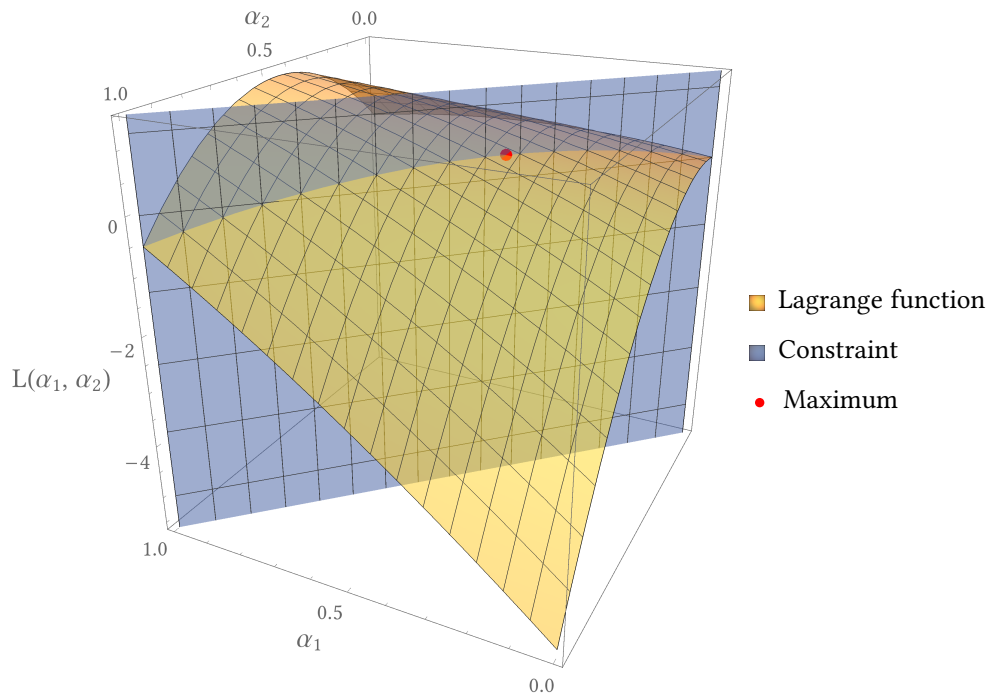
Figure 3: 3D plot of the Lagrange function $L(\alpha_1, \alpha_2)$ (Equation 1) in orange together with the constraint of Equation 2 in blue. Valid solutions are restricted to the intersection curve between the constraint plane and the Lagrange function. The point marked in red corresponds to the extrema value (which is a maximum in this case).
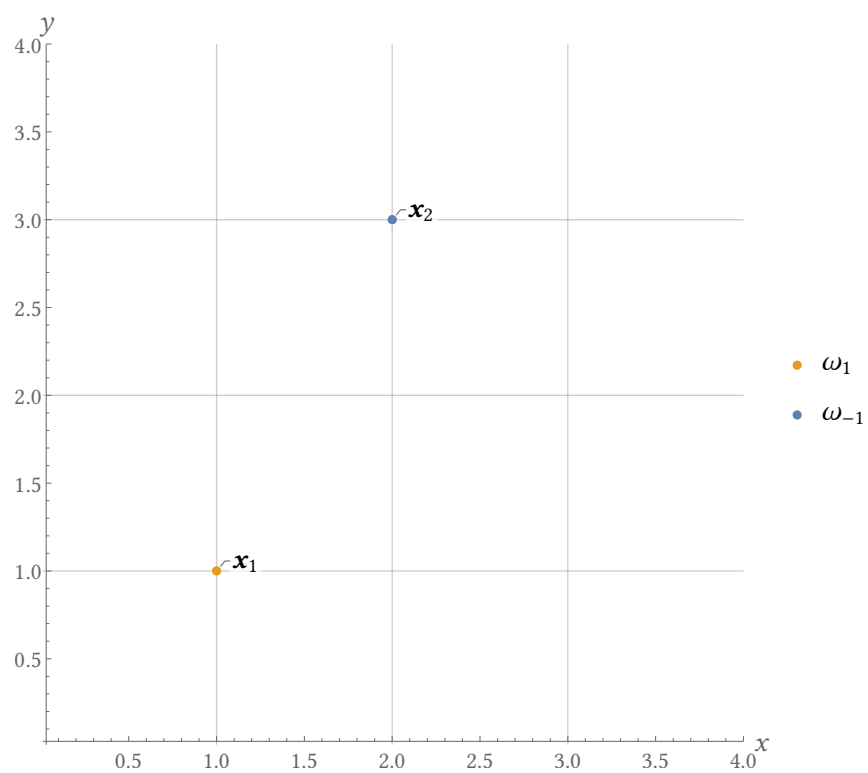
Figure 4: Two points from two classes corresponding to the problem of Exercise 2.