ML: Algo and Theory
SS 18
Tutor:

1	2	3	4	$\sum$

Florence Lopez (3878792)
Jennifer Them (3837649)

Assignment 5

(Due 29. April 2018)

Exercise 1

- see code.
- see code.
- see code.
- see code. $C = 100$ seems to work the best, since it produces the highest cv and test accuracy.

Exercise 2

- see code.
- see code. With increasing C the hyperplane gets less data dependant.
- see code. $C = 0.01$ seems to work best, since the test accuracy is the highest for this one.
- see code. Observation: The accuracy for X_1 and Y_1 is higher, since the data seems to be better separable as one can see in the plot. This is not the case for X_2 and Y_2, since the datapoints for the different classes are somehow overlapping and not easy to separate.

Exercise 3

(a)

We say a hyperplane, which is defined by w and b, is in canonical representation when $min_{i=1,..,n} |\langle w, x_i \rangle + b| = 1$ with respect to the data points is given. $Y_i(\omega^T X_i + b) \geq 1$ makes sure that every data point has at least the distance 1 from the hyperplane. When you find the optimal solution by minimizing ω, this imposes that there is at least one point (trainingsdata) with the distance 1. Which means that the minimal distance is 1.

(b)

The objective function $\frac{1}{2}||\omega||^2$ is strictly convex (see ANALYSIS OF SUPPORT VECTOR MACHINES by Shigeo Abe and Uniqueness of the SVM solution by Burges and Crisp). Because the training data is linearly separable there is a solution. Due to the strict convexity of the objective function there is only one point that represents the minimum so that the solution is therefore unique.

.