

Assignment 4

Computational Intelligence, SS2020

Team Members		
Last name	First name	Matriculation Number
Blöcher	Christian	01573246
Bürgener	Max	01531577

1 Linear SVM

1.1 Plots

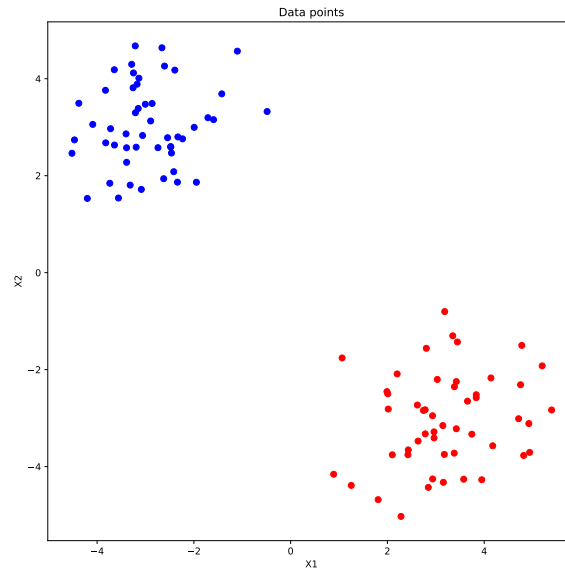
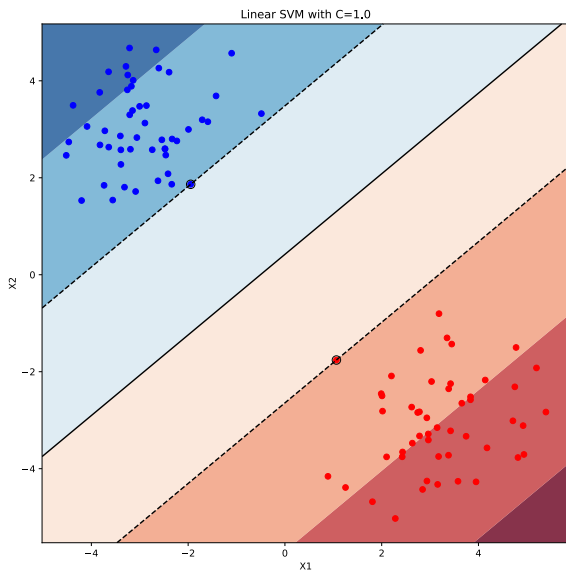
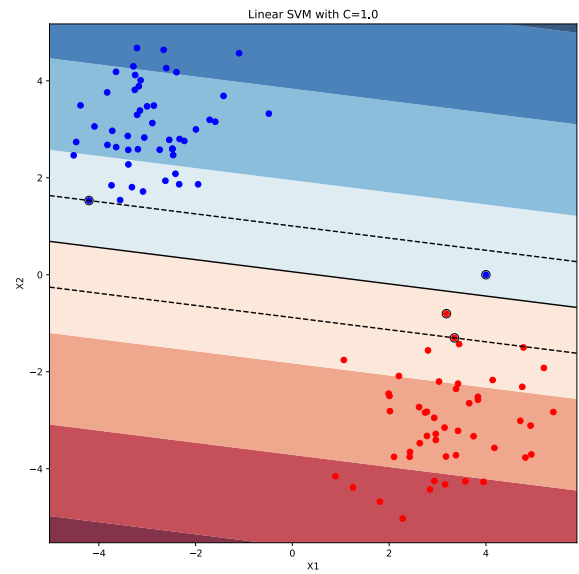


Figure 1: Dataset

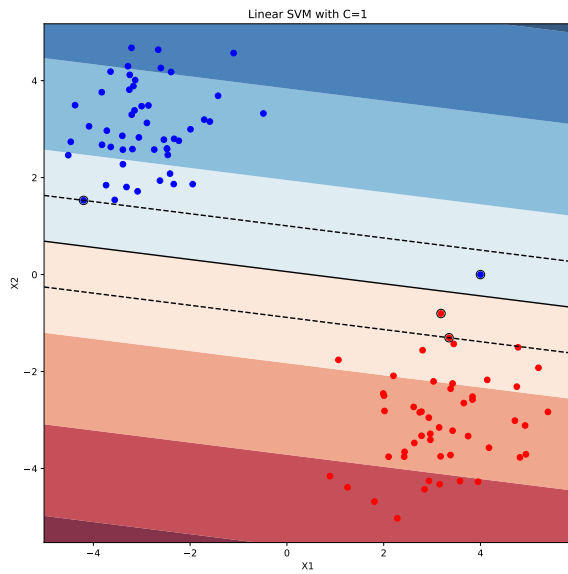


(a) Dataset

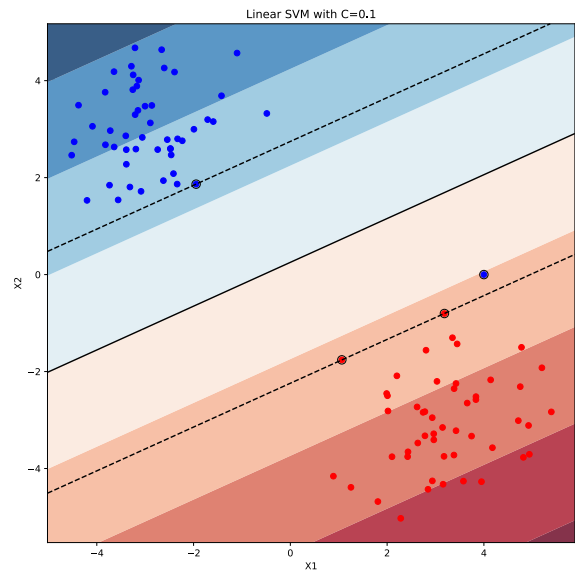


(b) Dataset with additional data point

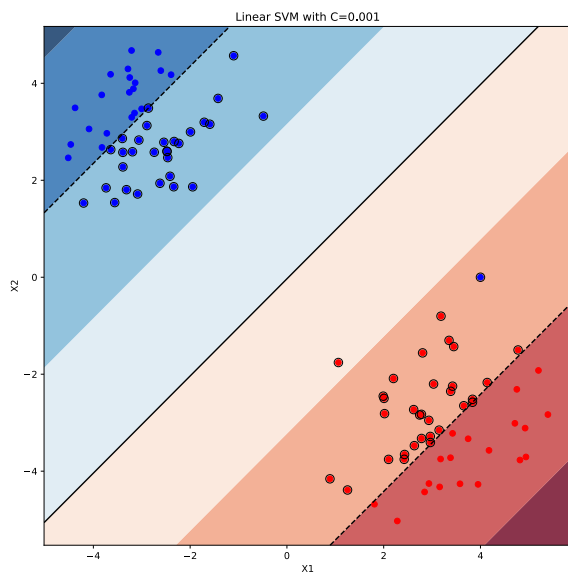
Figure 2: Classification of the dataset using SVM



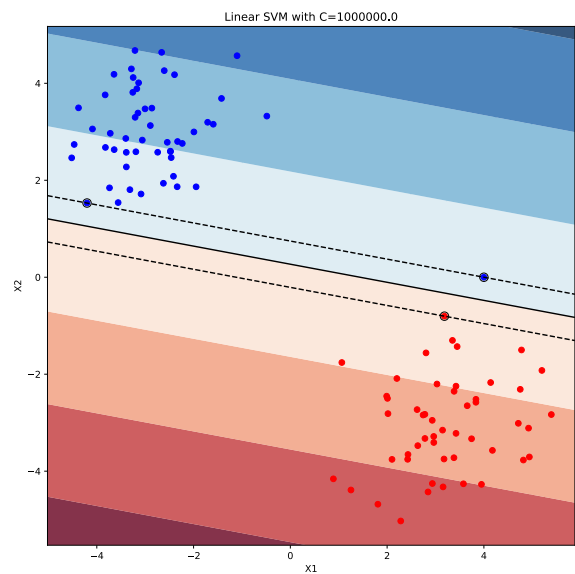
(a) $C = 1$



(b) $C = 0,1$



(c) $C = 0,001$



(d) $C = 1000000$

Figure 3: SVM Classification using different C parameters

1.2 How and why changed the decision boundary when the new point was added?

The margin is defined as the perpendicular distance between the decision boundary and the closest data point. Since the new added data point is located in the margin the algorithm adjusts its solution by choosing the option with the smallest generalization error. Apparently this solution includes more support vectors (data points within and on the margin boundary).

1.3 The inverse regularization parameter C

- C controls the penalization of data points within the margin. It is scaling data samples so that samples can be misclassified or lie within the margin boundary. This is helpful since not all problems can be clearly separated.
- Figure 3 is showing the classification using different values for C . The margin boundary gets larger with decreasing values for C . As a result the classification is showing more support vectors and also misclassifies the additional data point since these points are less penalized. A very large value for C leads to the strict constraint that the margin borders are denoting the closest data point to the decision boundary. Therefore the margin boundary is very small.

2 Nonlinear (kernel) SVM

2.1

2.2

2.3

2.4

2.5

3 Multiclass classification

3.1 'One vs. Rest' vs. 'One vs. All' algorithms

SVMs are fundamentally two-class classifier. We can use different approaches to solve problems with K classes.

The 'One vs. Rest' algorithm compares the to be determined class (positive) with the rest of the dataset (negative). During the scoring phase the algorithm determines the probability for each class of a sample using the binary classifiers.

'One vs. All' creates $K(K - 1)/2$ classifiers, which means that we use all possible pairs of classes. Each binary classifiers predicts one class. For classification the class with the highest number of predictions will be chosen. That algorithm requires in comparison to the 'One vs. Rest approach' much more time to get trained and to classify.

3.2 Classification using linear and RBF

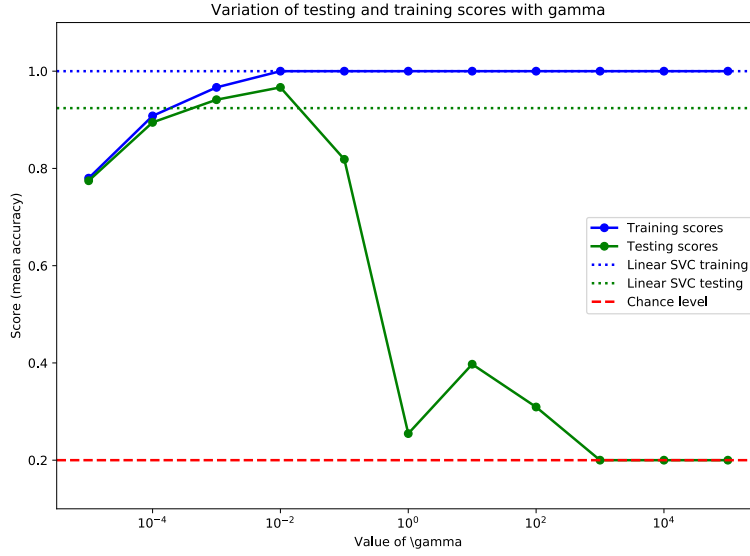


Figure 4: Scores using linear and RBF kernel on MNIST Dataset

- Figure 4 shows that the linear kernel performs very well on MNIST image classification. Using $\gamma = 0,1$ the RBF classification performs slightly better than the linear kernel. For higher γ values the classification overfits and the testing scores are quickly descending.
- The linear kernel performs good, because the sample features which are generated from the MNIST pictures must be very good linear separable in higher dimensions.

3.3 Confusion matrix

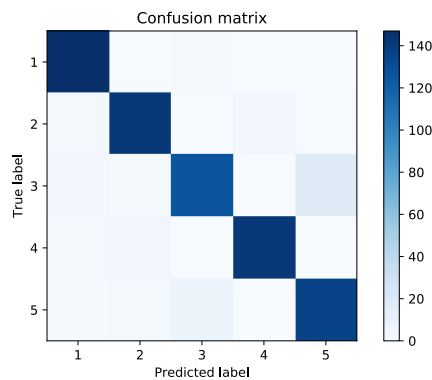


Figure 5: Confusion matrix

	1	2	3	4	5
1	147	1	2	0	0
2	3	142	0	4	1
3	4	2	126	0	18
4	3	5	0	142	0
5	2	3	9	0	136

Table 1: Confusion matrix

The rows of our confusion matrix correspond to the true classes and the columns to the predicted classes. Class three is the most misclassified class with score of 84%.

3.4 Misclassified images



Figure 6: Misclassified images

Figure 6 shows the column of class three. The images are showing the real numbers, which were classified as class three. Apparently the algorithms has problems to distinguish threes and fives. Probably because the lower part of these numbers are very similar. The reason for both falsely determined ones could be that they are untypical ones. The first one is very bold and the other is a bit rotated.

4 SVM with Gradient Descent

4.1

4.2

4.3

4.4

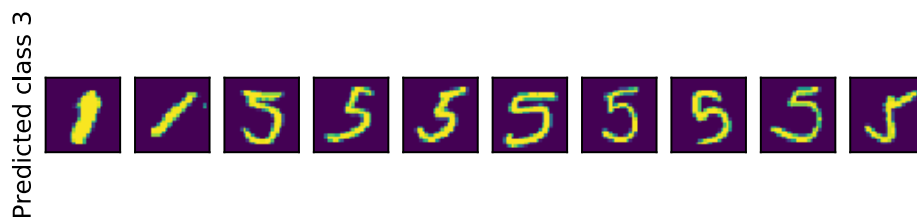


Figure 6: Misclassified images

Figure 6 shows the column of class three. The images are showing the real numbers, which were classified as class three. Apparently the algorithm has problems to distinguish threes and fives. Probably because the lower part of these numbers are very similar. The reason for both falsely determined ones could be that they are untypical ones. The first one is very bold and the other is a bit rotated.

4 SVM with Gradient Descent

4.1

4.2

4.3

4.4