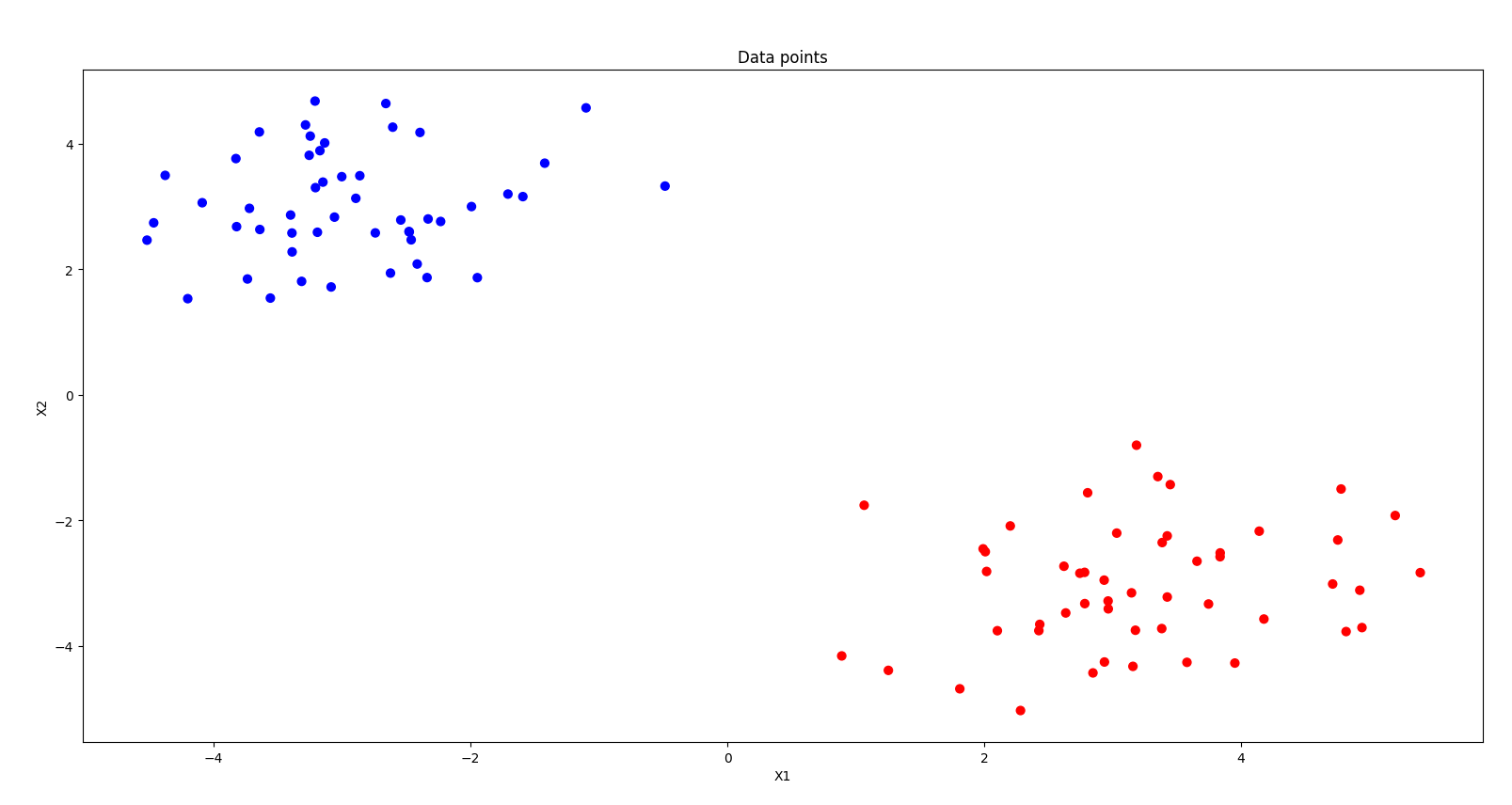
**Assignment 3**

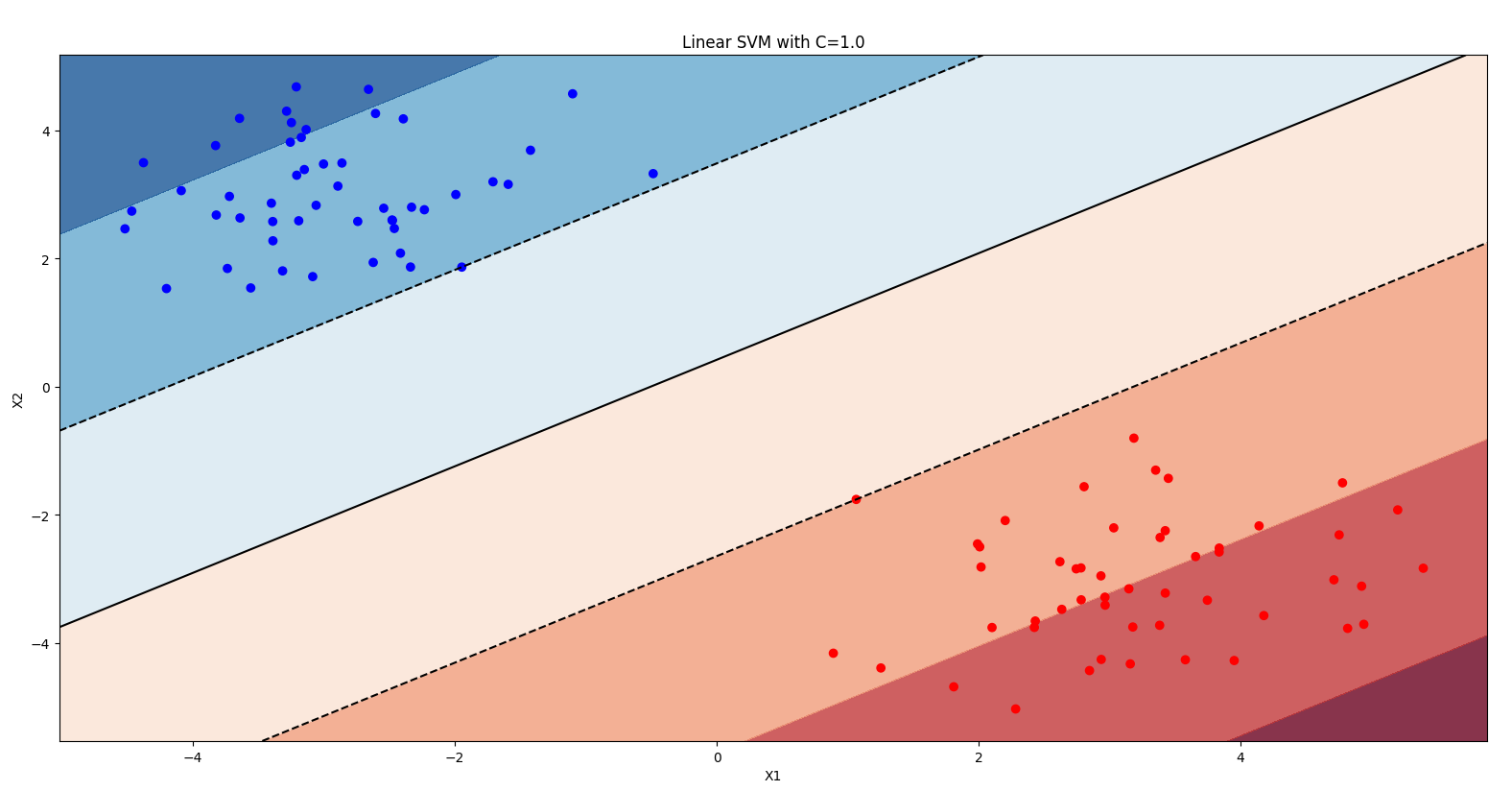
# Computational Intelligence SEW, SS2017

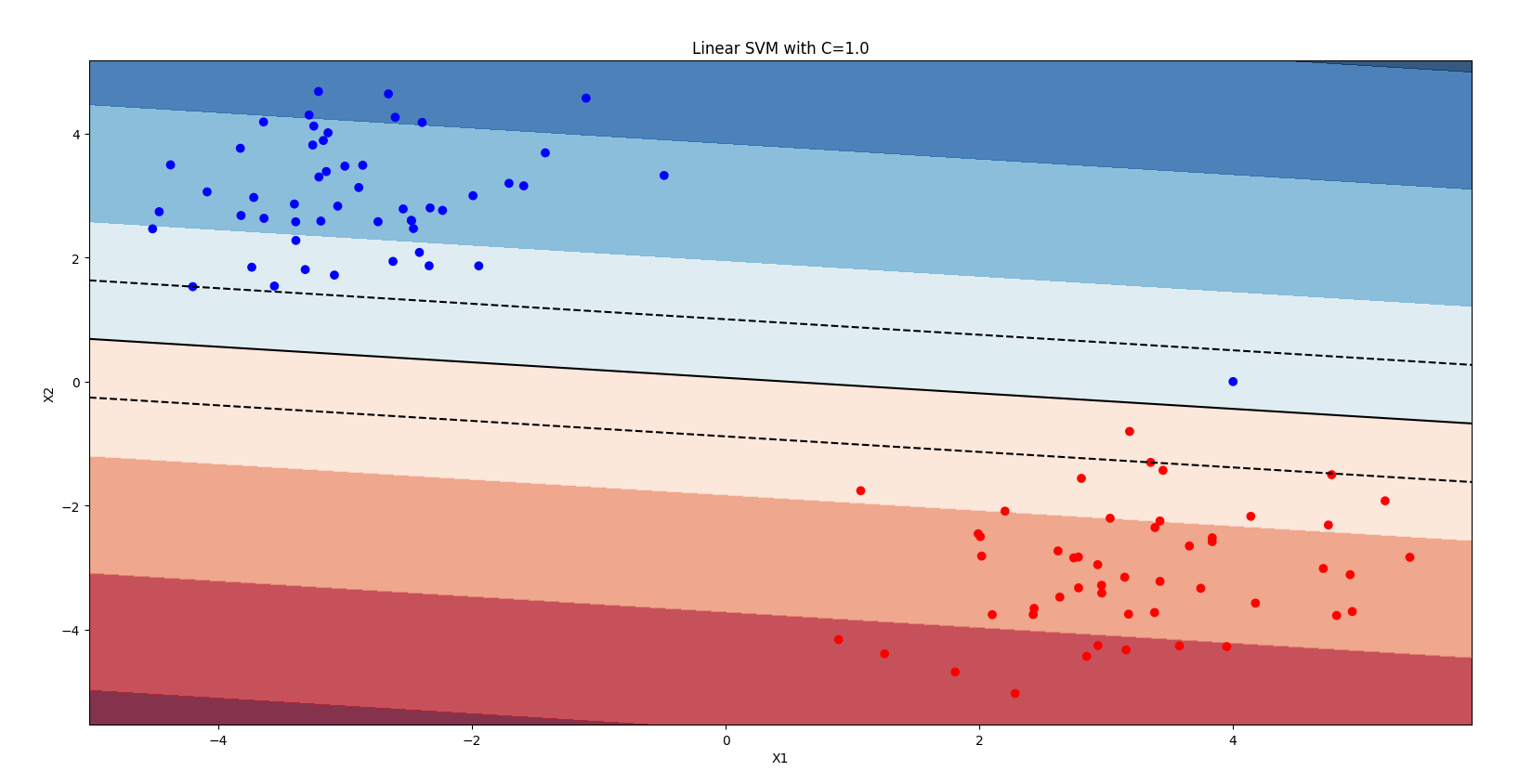
|  |  |  |
| --- | --- | --- |
| **Team Members** | | |
| Last name | First name | Matriculation Number |
| Papst | Stefan | 1430868 |
| Guggi | Simon | 1430534 |
| Perkonigg | Michelle | 1430153 |

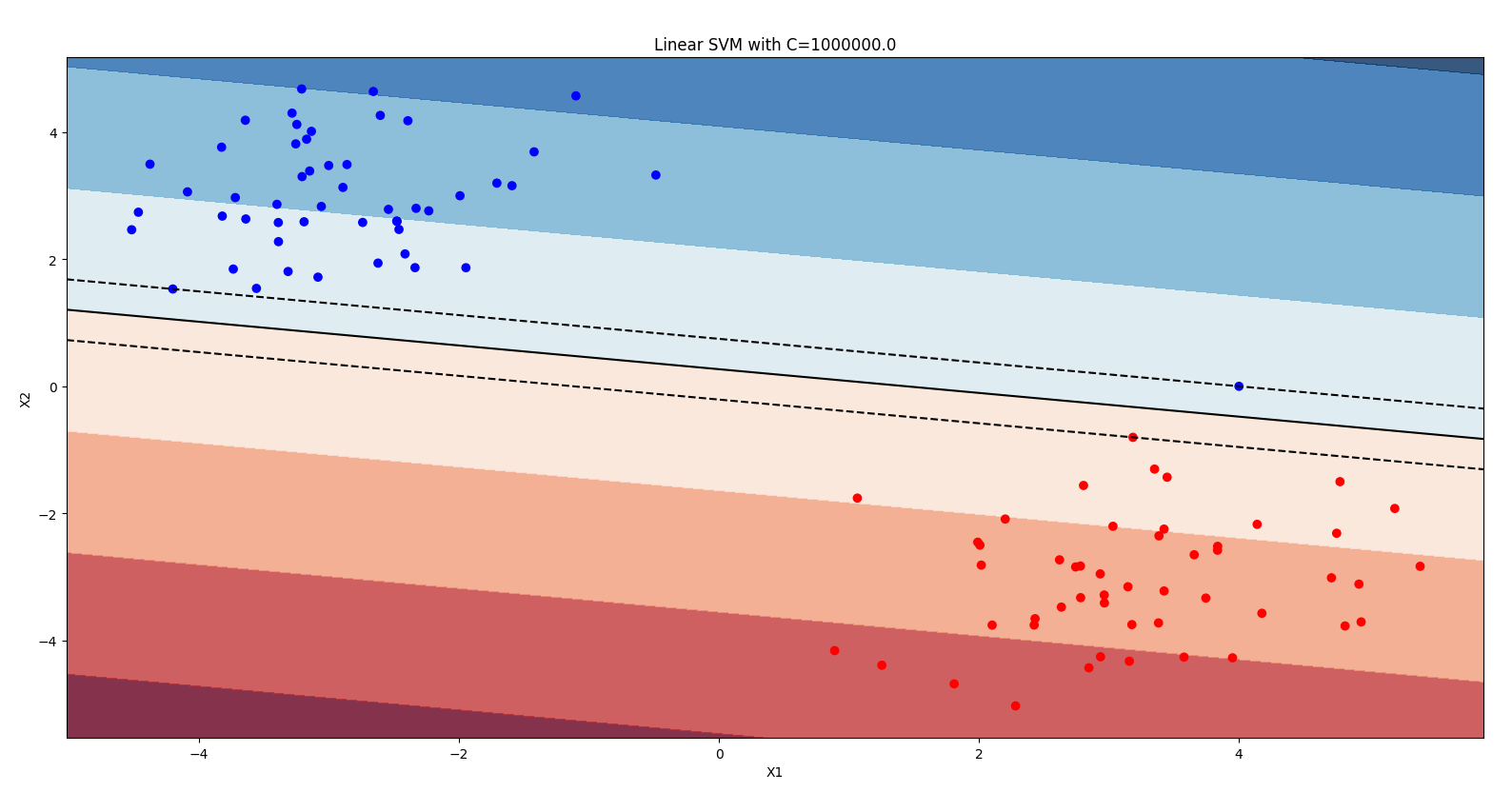
**1 Linear SVM**

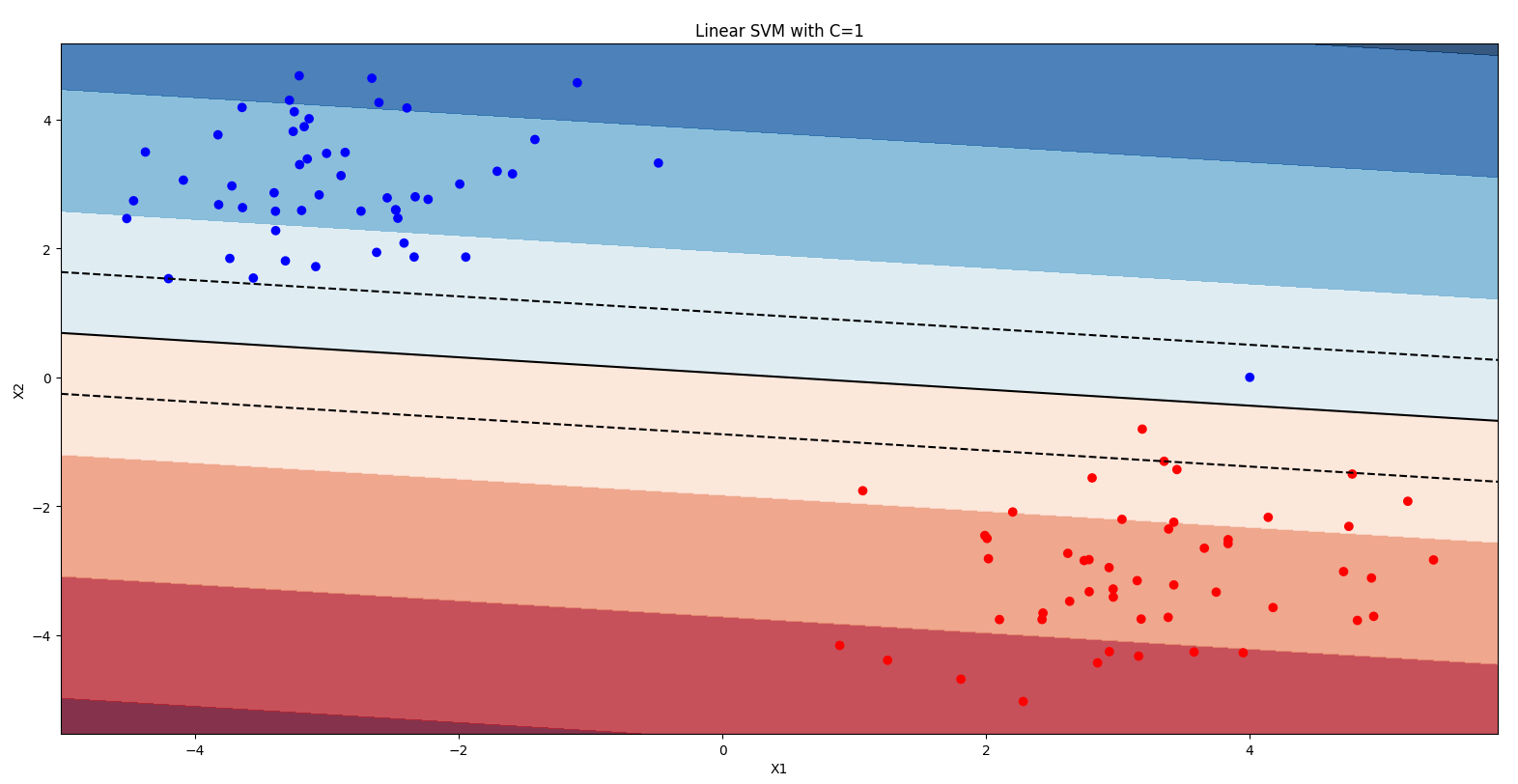
***Include plots of all the results***

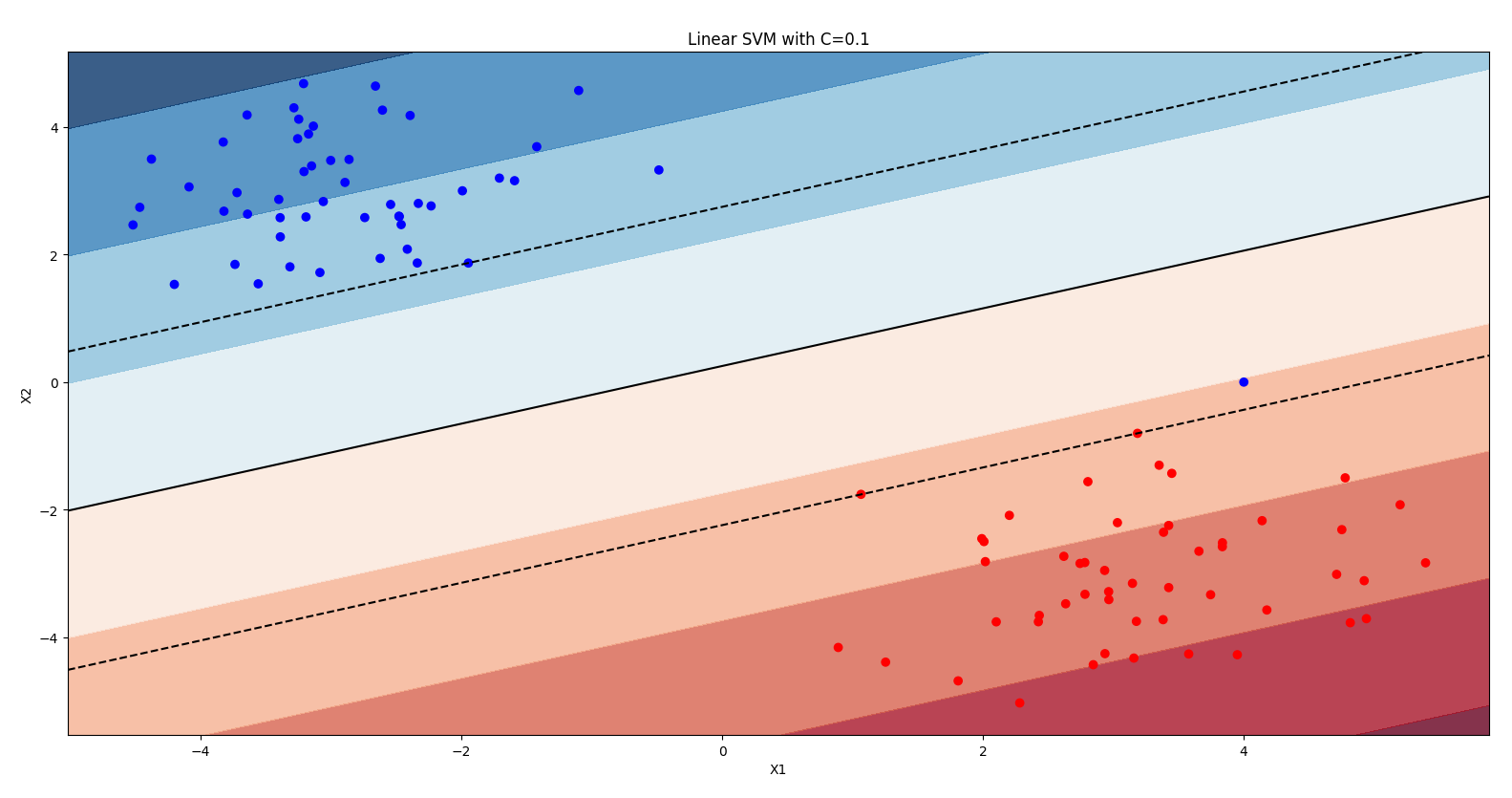


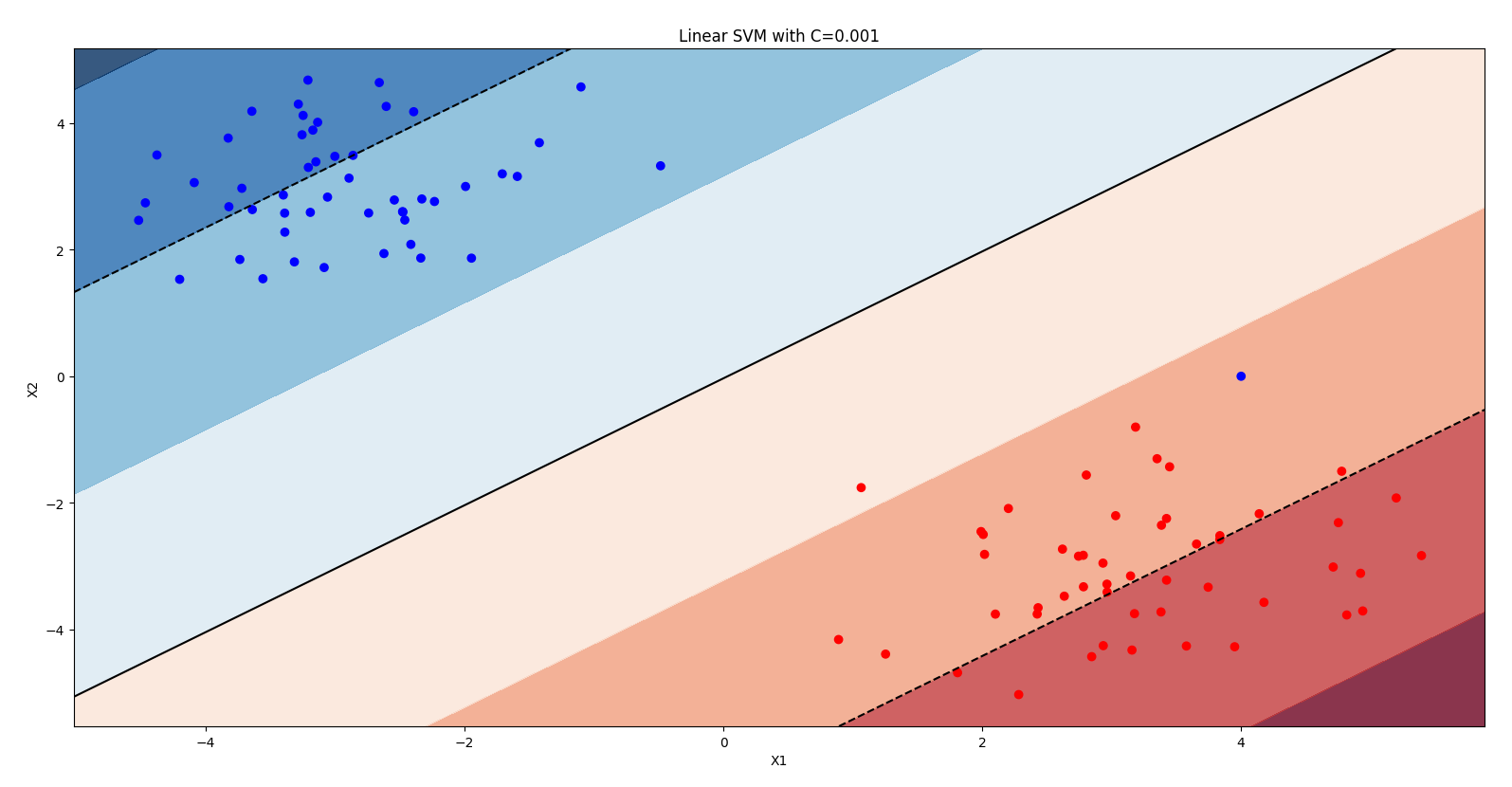


Task b: 

Task c:







***For task b), discuss how and why the decision boundary changed when the new point was added.***

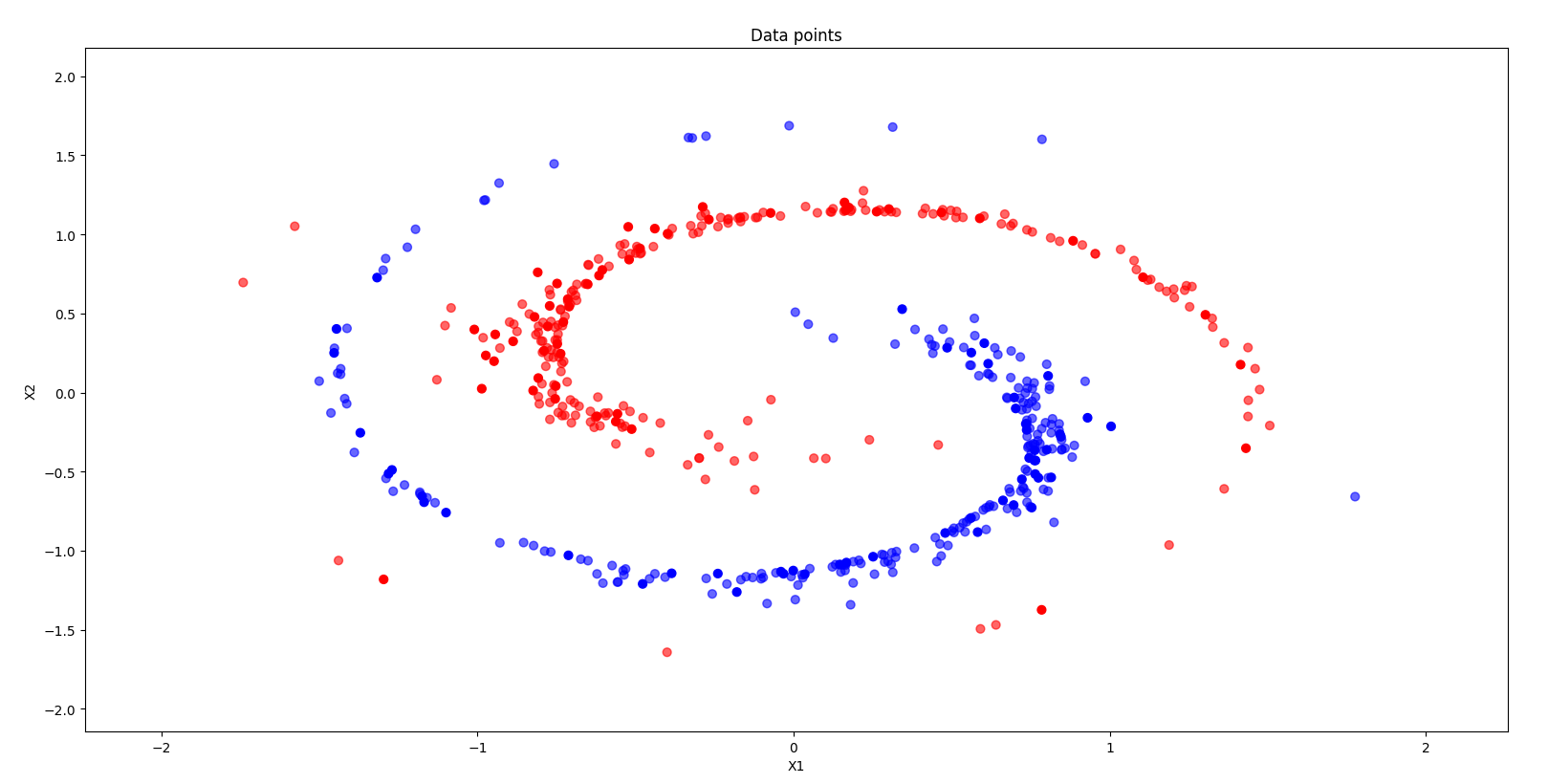
The SVM tries to find a decision boundary that separates the classes. When adding the new point, the blue point, on the specific position (4, 0) that is very isolated from the rest of the points of the blue class the decision boundary must be set new including this significant point. This can be seen at the third plot.

***For task c):***

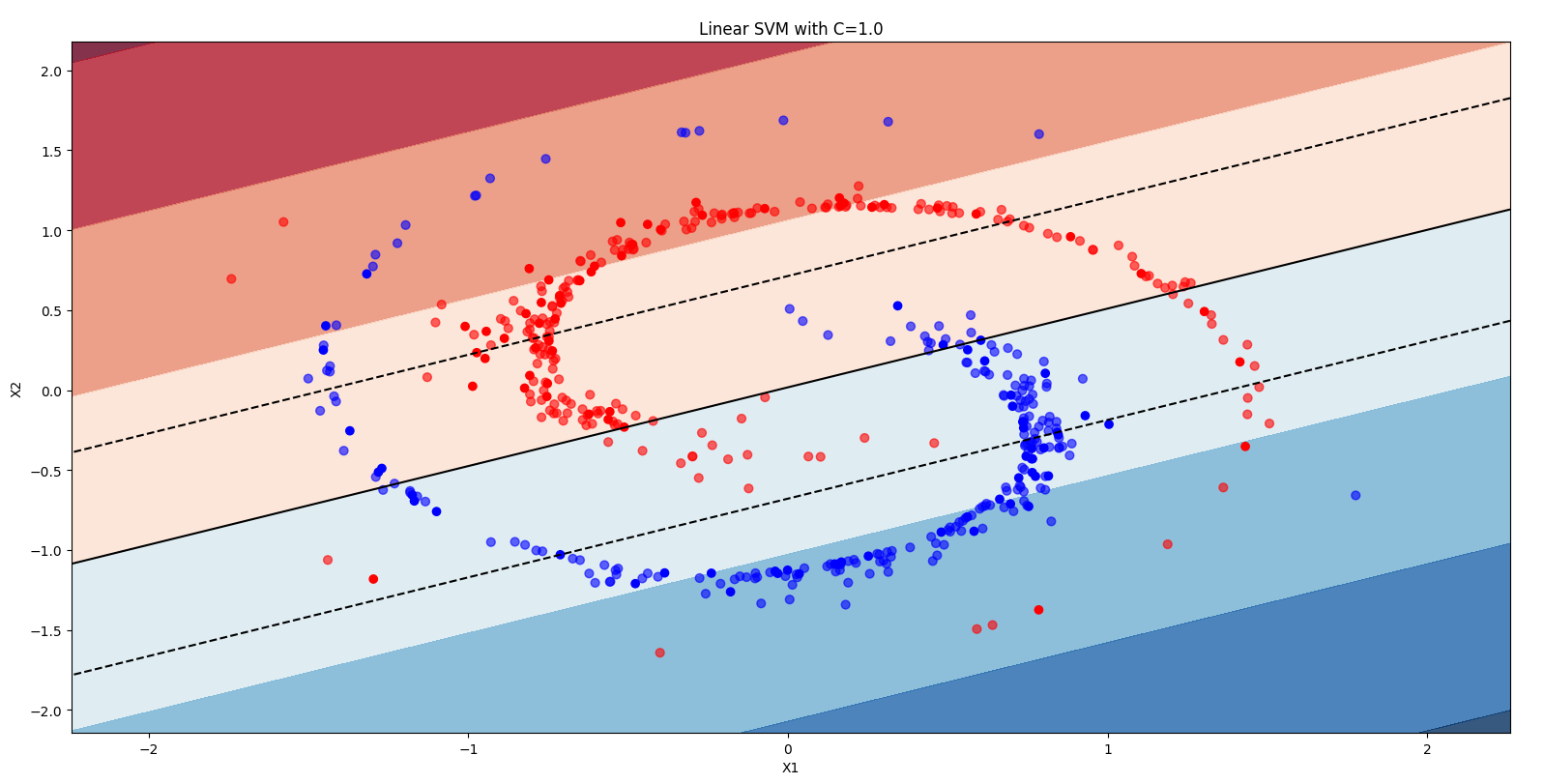
* ***Report how the parameter C influences the decision boundary found by the SVM***  
  The parameter *C* determines whether you want to find a hyperplane with the largest minimum margin or a hyperplane that separates the samples correctly (as many as possible). Thus, we can say that the parameter *C* describes the sensitivity of the SVM to outliers like the new point added. A high *C* represents high sensitivity and (almost) correct classification, whereas a low *C* represents low sensitivity and the possibility for misclassification.
* ***How does the number of support vectors found by the SVM change with the value of C? Why?***

**2 Nonlinear (kernel) SVM**

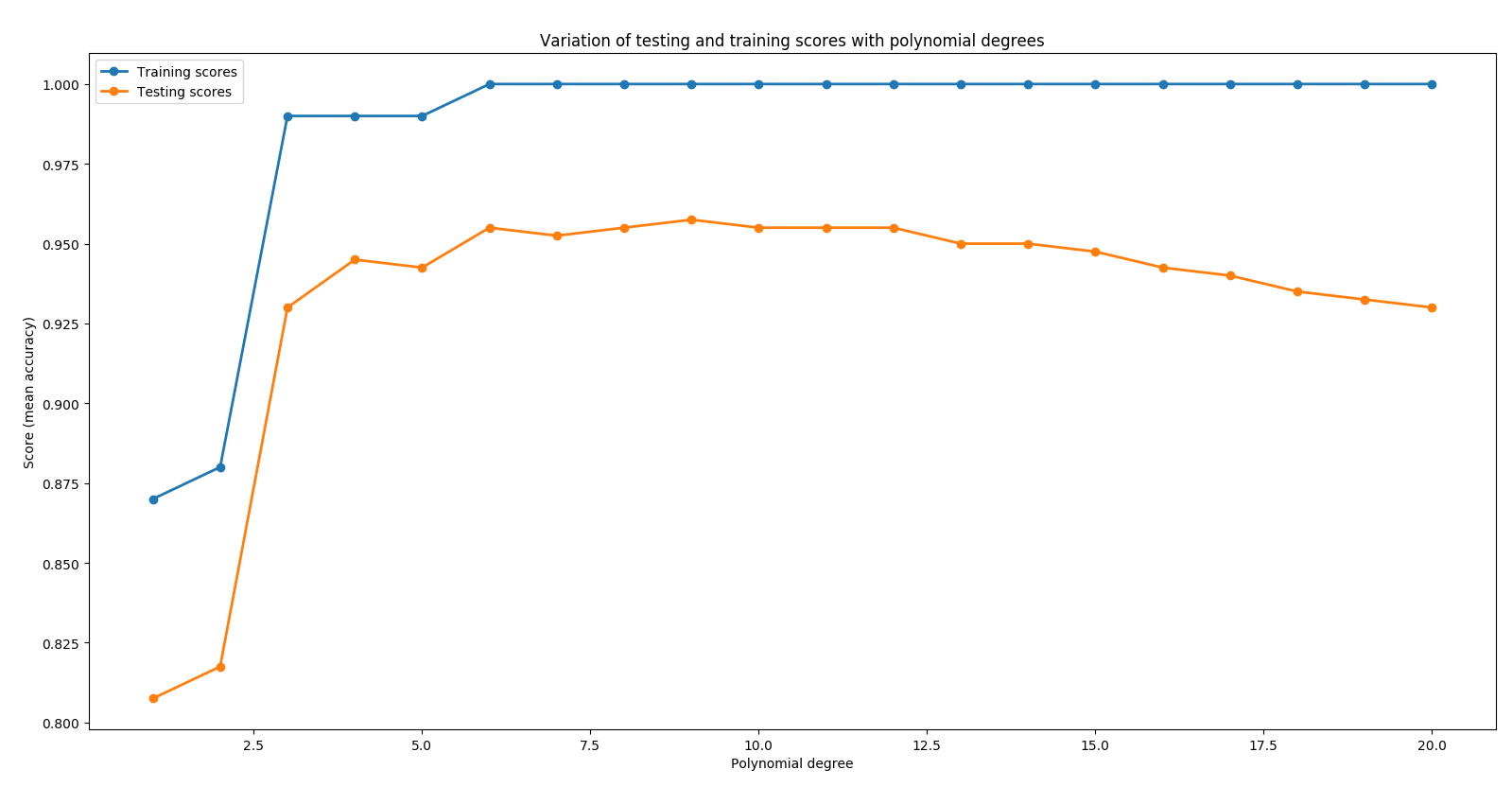
***Include plots of all the results***

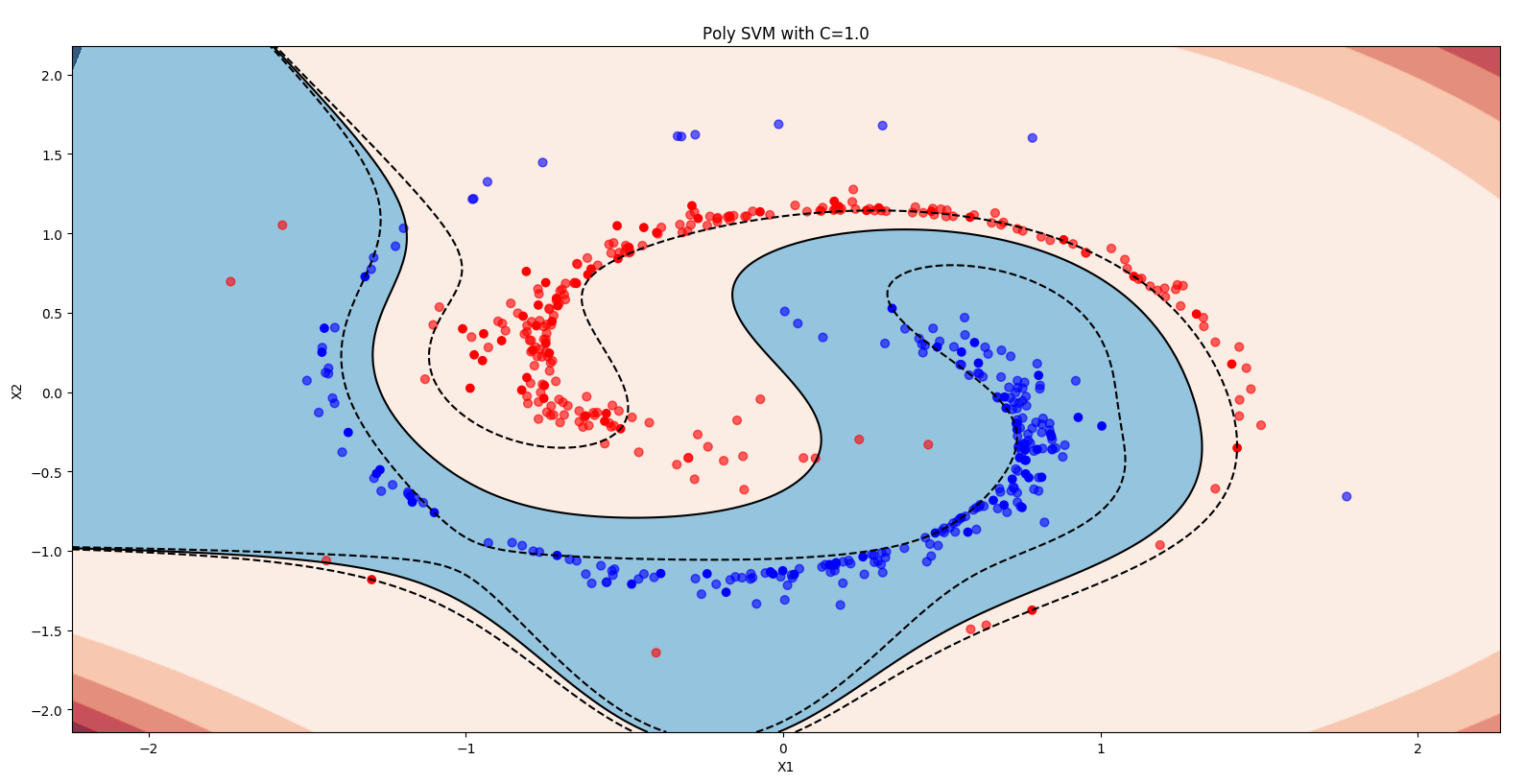


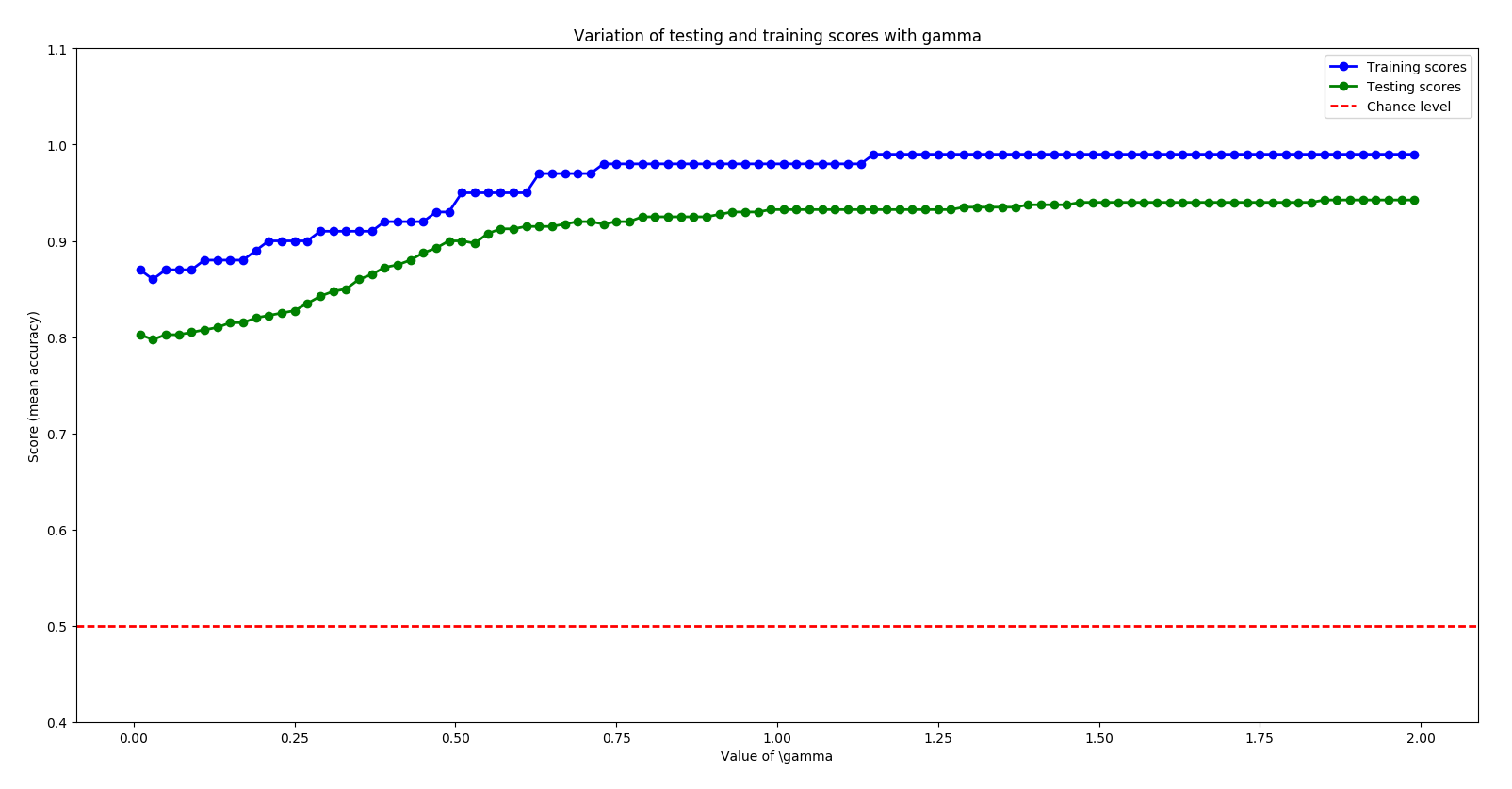
Task a:

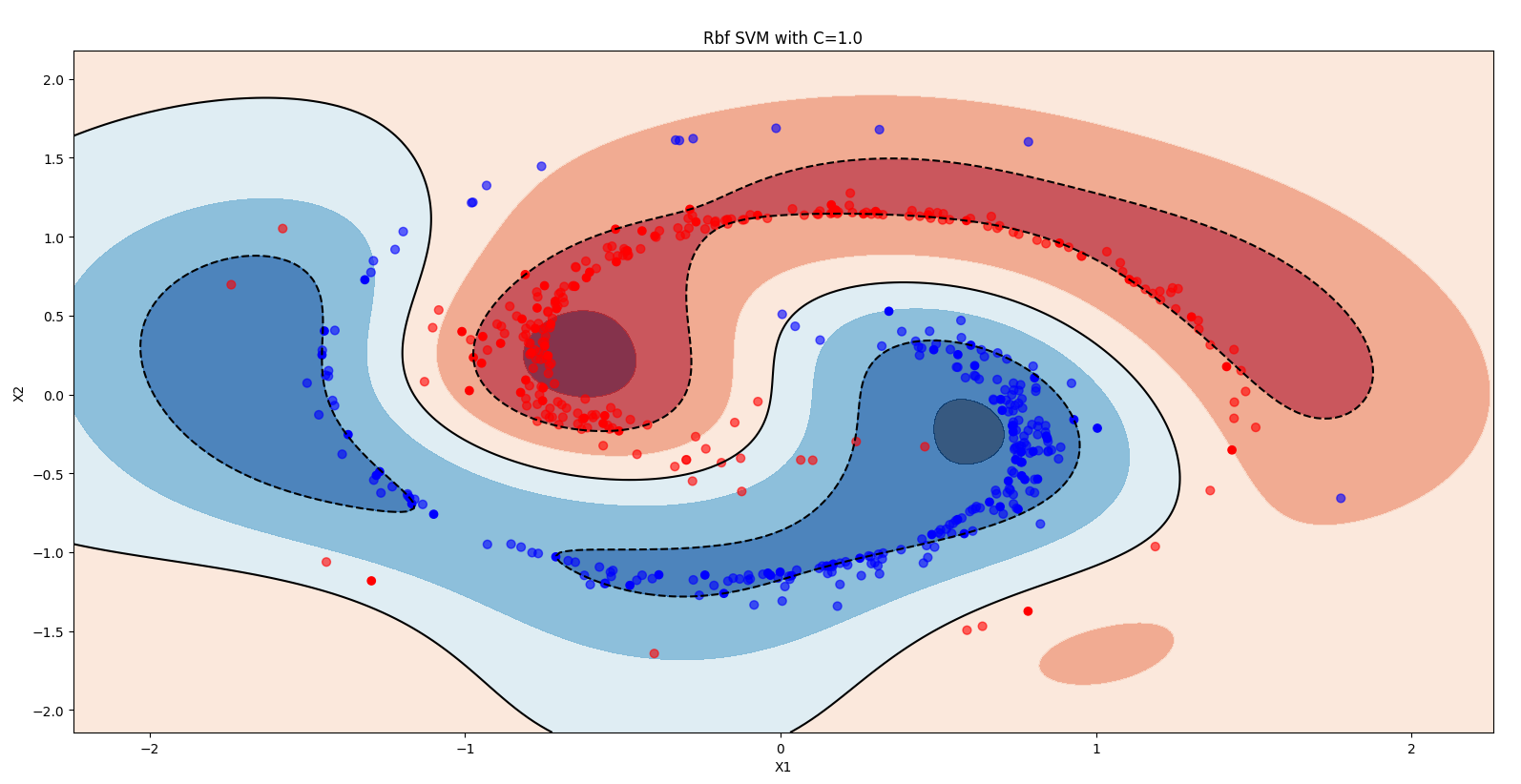


Task a:

Task b: 



Task c: 



***For task b), which degree of the polynomial produces the highest test score (accuracy)? Report this test score.***

degree for highest test score: 9

highest test score for poly kernel: 0.9575

***For task c), which value of gamma produces the highest test score (accuracy)? Report this test score.***gamma of highest test score: 1.85  
highest test score for rbf kernel: 0.9425

***Compare results obtained by each of these three kernels:***

* ***State the maximum test score achieved for each of these kernels and the kernel parameter for which that was achieved.***test score for **linear kernel** 0.8125  
  highest test score for **poly kernel** 0.9575 with degree 9   
  highest test score for **rbf kernel** 0.9425 with gamma 1.85
* ***Which of the considered kernels performs best and why?***
* ***Compare the complexity of decision boundaries and the number of Support Vectors found.***
* ***Which kernel generalizes best for the given dataset?***

**3 Multiclass classification**

***Recall the algorithms `One-versus-Rest' (or versus-all) and `One-versus-one' multi-class classification procedures. How many binary classifiers need to be trained in both cases?***

One-versus-Rest: N classifiers for N classes (one per class)

One-versus-One: N (N-1) / 2 classifiers for N classes

***Include plots for ex\_3\_a with the scores of a linear and a rbf kernels.***

***Discuss those results. In particular why does a linear kernel perform well on images?***

***Find the digit class for which you get the highest error rate.***

***Include plots for ex\_3\_b of the confusion matrix and the first 10 images from the test set of the most misclassified digit.***

***With the help of these two plots, discuss why the classifier is doing these mistakes.***