# Task 1

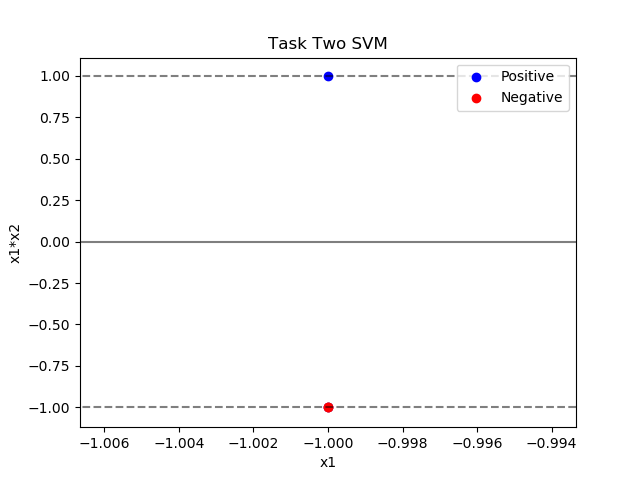
Q1: Margin and support vectors are the bounds used to separate classes in a SVM. The margin is the space between two or more classes that serves as a buffer zone for hard classifications (in the case of soft margins some data points can exists in this margin). The support vectors are the outer bounds of the margin. Often, they touch the inner extremes of two or more classes such that they form a dividing line (border) between classes.

Q2: SVM can deal with non-separable data by either forming a non-linear bound around classes. As well as a SVM can translate the data into the next highest dimension.

Q3-4: A kernel is a function that can find the pair-wise distance between two feature sets without having to directly translate the data into a higher dimension.

# Task 2

See code for more details…

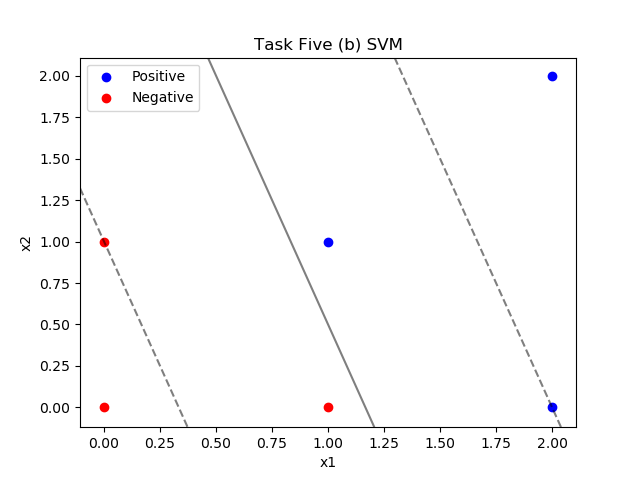


# Task 3

(x1-a)2 + (x2 – b)2 – r2 = 0

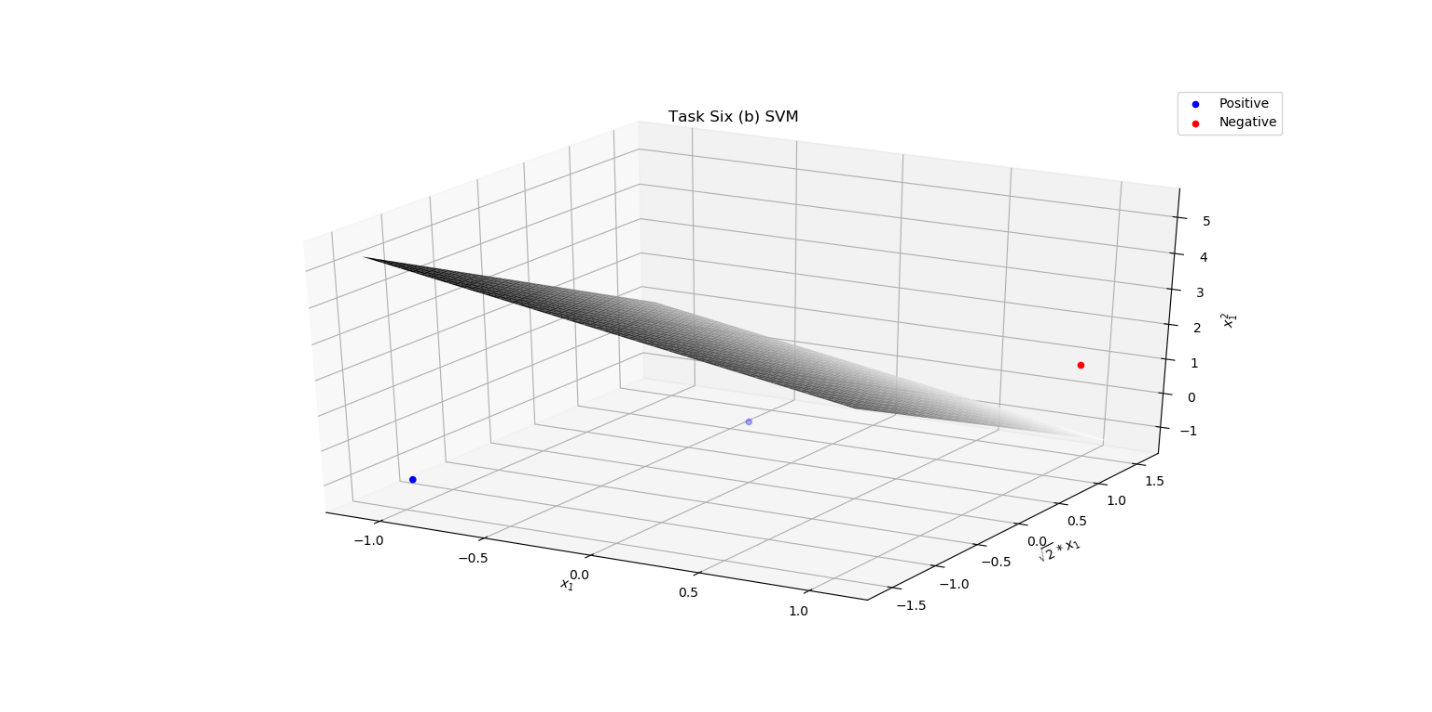
# Task 4

# Task 5

Yes, they are linearly separable. 

# Task 6

1. No they are not linearly separable because the positive class is a straddle point between the two negative classes.



Task 7

|  |  |
| --- | --- |
| Kernel | Average Accuracy (5 Fold Cross-validation) |
| Linear | 78.68% |
| Quadratic | 80.14% |
| RBF | 80.36% |

Additional Questions:

• Approximately how many hours did you spend on this assignment?

20 hours

• Which aspects of this assignment did you find most challenging? Were there any significant stumbling blocks?

The kernel conversion math. It was not taught really at all during class and this is something that needs some examples rather than just “it exists for this purpose”.

• Which aspects of this assignment did you like? Is there anything you would have changed?

Liked: Simple/small data sets

Disliked: The titanic datasets, it has been used for every hw assignment and at least the “original” version of this homework assignment had a different dataset first to mess around with (but still had titanic).