

COMS 474  
Homework 7

Haadi Majeed

Spring 2022

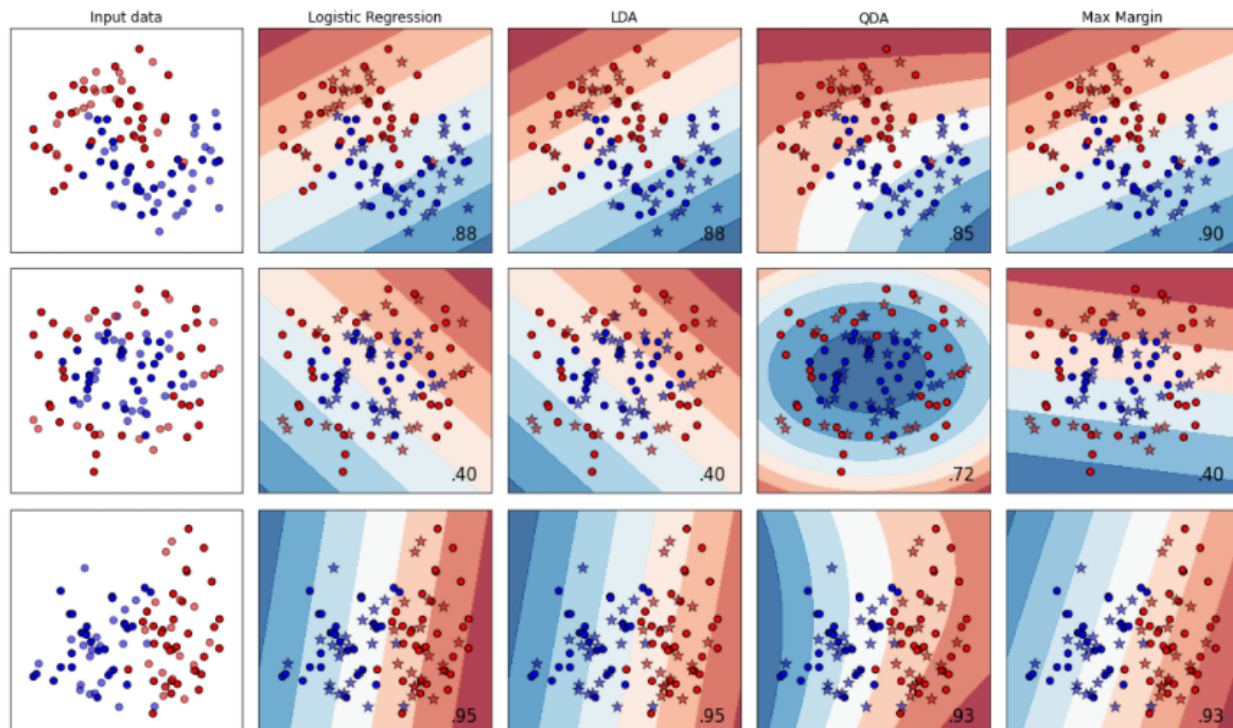
## Problem 1

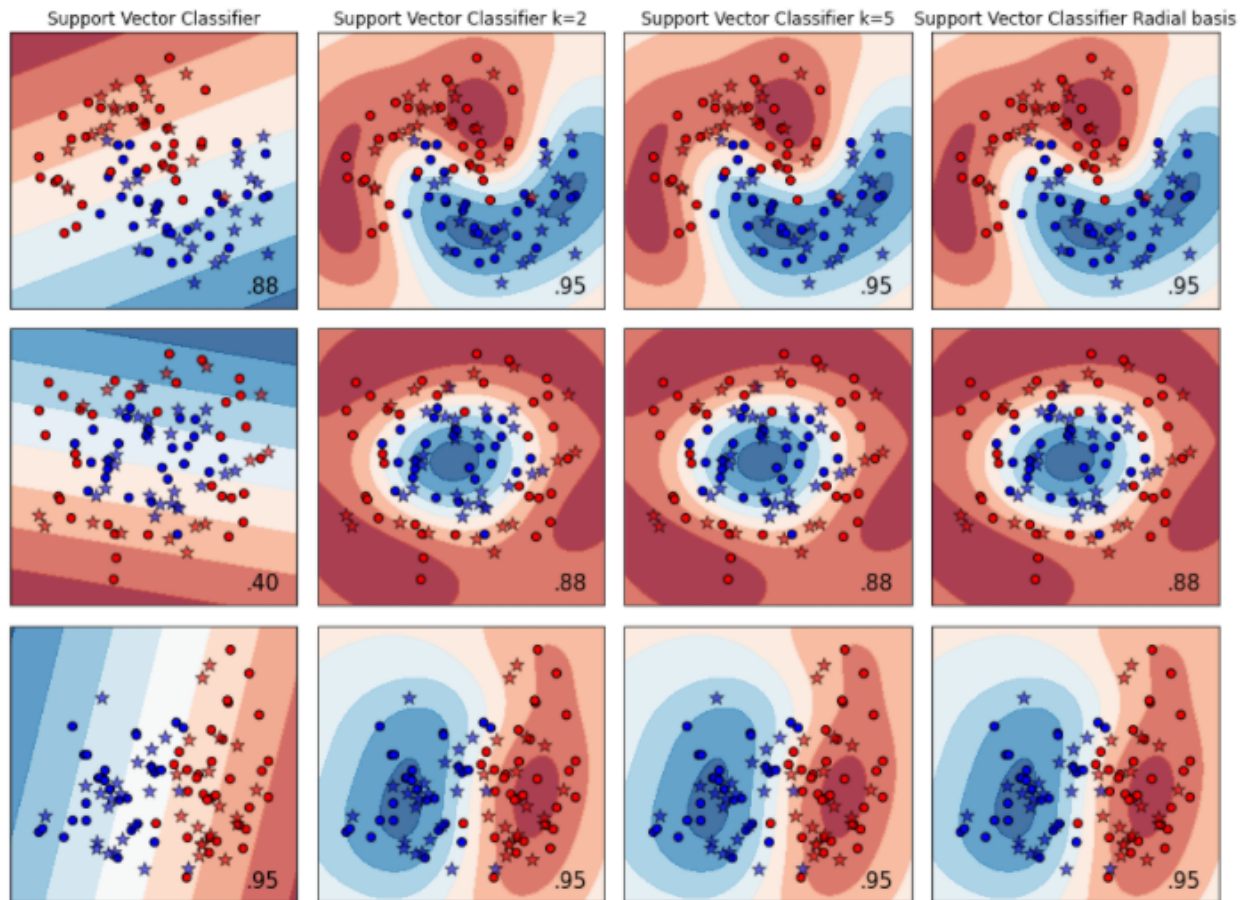
[30 points]

For this problem you will fit classifiers to three simple data-sets and visually examine the classification boundaries

**A**

Copy of Plot





## B

Based on discussions in lecture, which methods do you expect to fit models that would result in linear decision boundaries? Was that the case? Briefly describe whether the methods that have linear boundaries appear to find the same/similar/different boundaries.

We expect that the following would have linear decision boundaries: Logistic Regression, LDA, Max Margin, SVC Linear. They mostly appear to have similar boundaries except for SVC Linear, which has the colours set differently, meaning that the predicted values would be inverted in comparison to the other models.

## C

Repeat the same for methods known to have boundaries defined by quadratic equations. QDA and the other SVCs have quadratic boundaries, which was expected behaviour since they are meant for quadratic analysis. The predictions are a bit more precise in some cases than others, but it is all as the sample data and test data see fit.

**D**

For methods whose boundaries can be non-linear, do they appear linear for the "linearly separable" data set?

Yes, it appears to look more linear-ish since the data it is fitting follows such a trend. Boundaries that are low order can be visibly seen in the models.

**E**

How do LDA and QDA compare for these data sets?

They are fairly close in this case, with QDA being a bit more flexible to the data, however in this case, that is a bad thing as QDA had a score of .93 vs LDA's .95 meaning LDA did a little bit better than QDA in this instance.

**F**

How do the different support vector methods compare?

The linear SVC scored lower for the first two data types than the quadratic SVCs except for the linearly separable data set, where the scores were equal for all 4 SVCs. For the second data set the linear SVC did significantly worse than the quadratic ones since it severely lacked the flexibility needed to fit the data properly.