Intro to ML - Exercise 4
Naama Avni 208523456

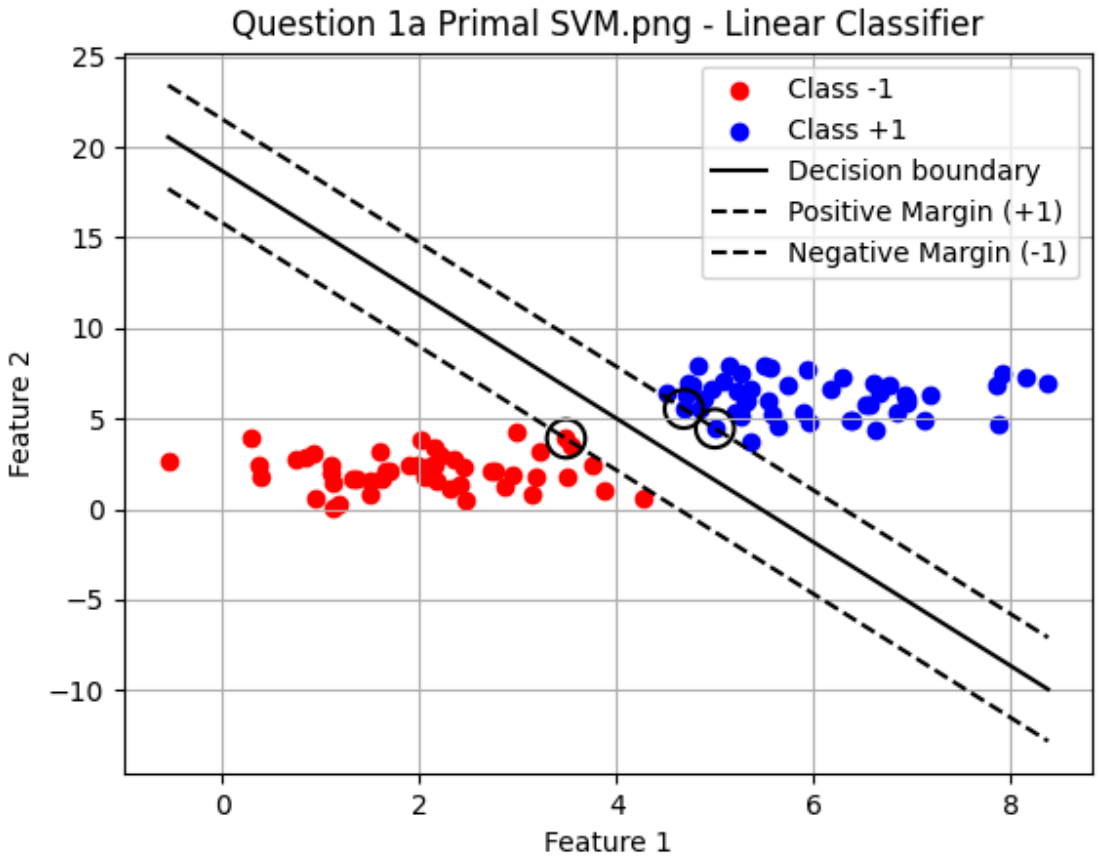
Outputs and solutions

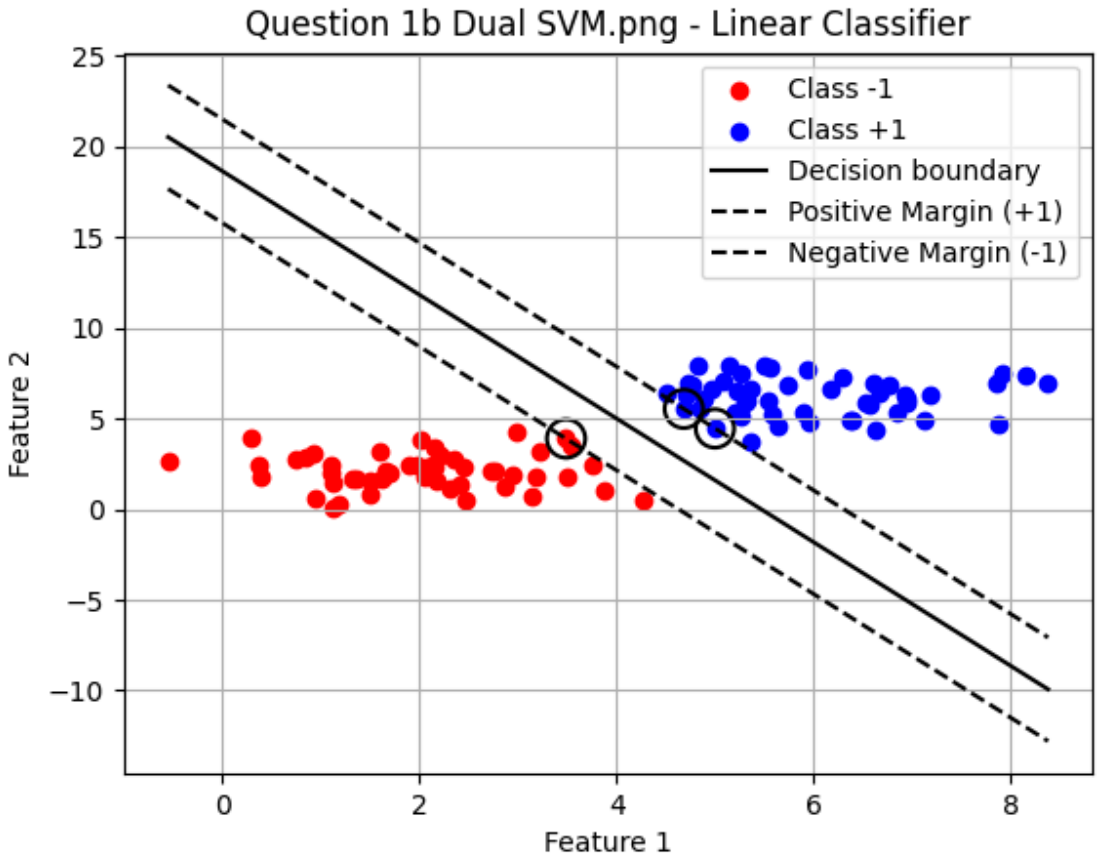
Question 1

Output:

QP solution: w = [1.1928781 0.3493699 -6.52192702]
Support vector indices: [35 42 46]
Translating to the dual we get: w_dual = [1.19220721 0.34978844 -6.52075124]
Support vector indices: [35 42 46]

Output plots:





Explanation:

As we can see the solutions we got from both technics the primal and dual SVM are identical as expected.
The w is [1.1928781 0.3493699 -6.52192702] and the Support vector indices are [35 42 46] in both cases.

Question 2

Output

Dataset split into:
Training set: 240 samples
Test set: 60 samples

Polynomial Kernel Results:
Degree | Train Accuracy | Test Accuracy

2	No solution found	
3	1.000	1.000
4	1.000	0.967
5	1.000	0.983
6	1.000	0.967

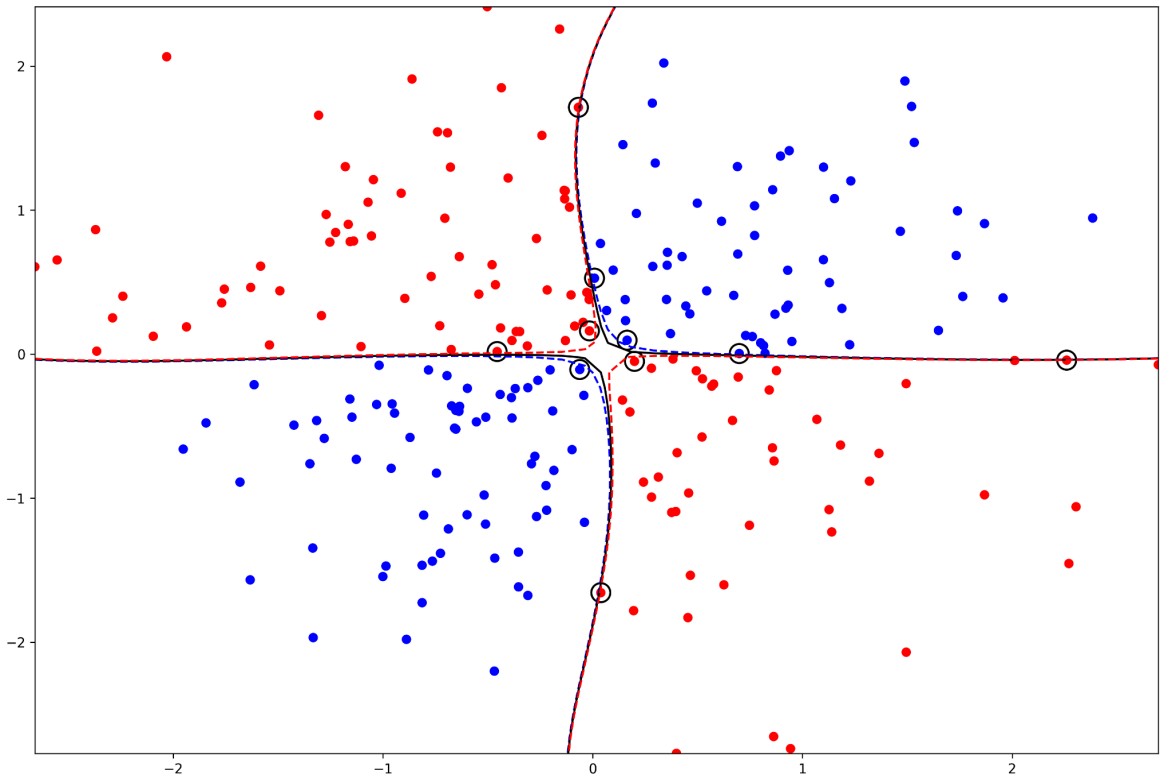
RBF Kernel Results:

Gamma	Train Accuracy	Test Accuracy
0.1	1.000	0.967
0.5	1.000	0.967
1.0	1.000	0.983
2.0	1.000	0.983
5.0	1.000	0.950

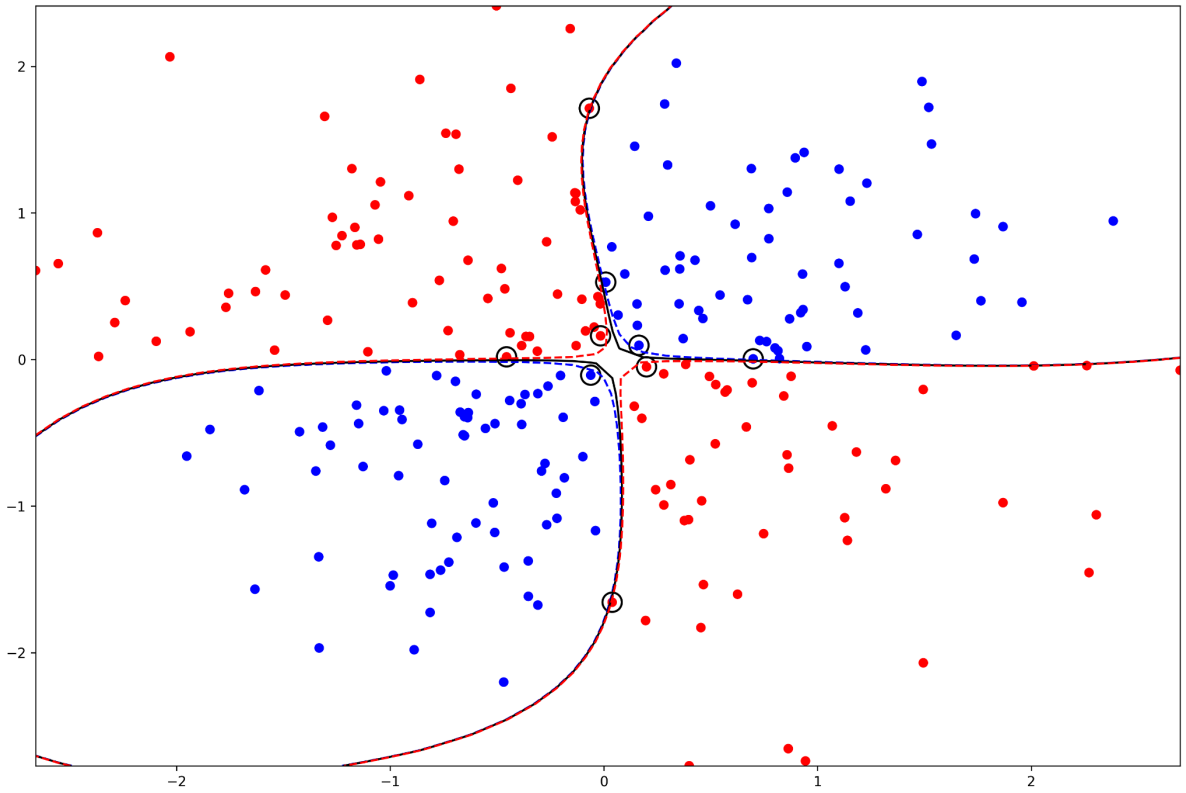
Output plots:

Selected models classification plots with the Support vectors(plots for all parameters available in the zip)

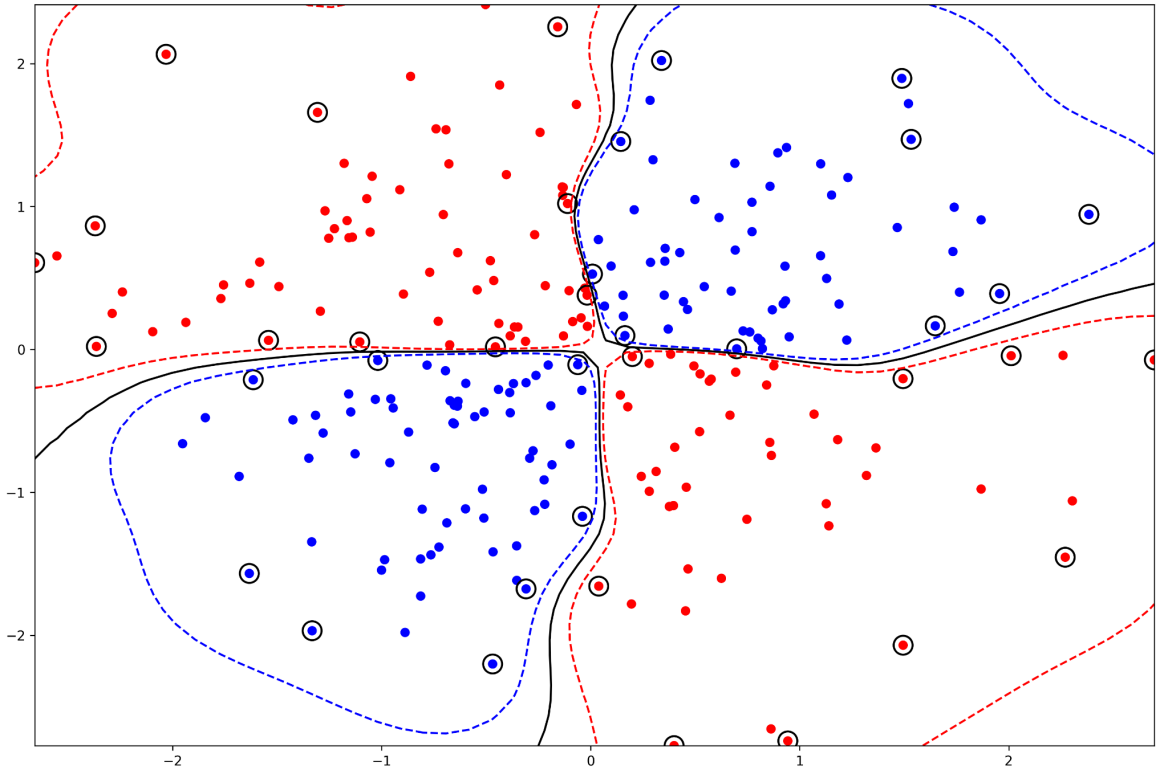
Dual SVM with polynomial kernel degree 5:



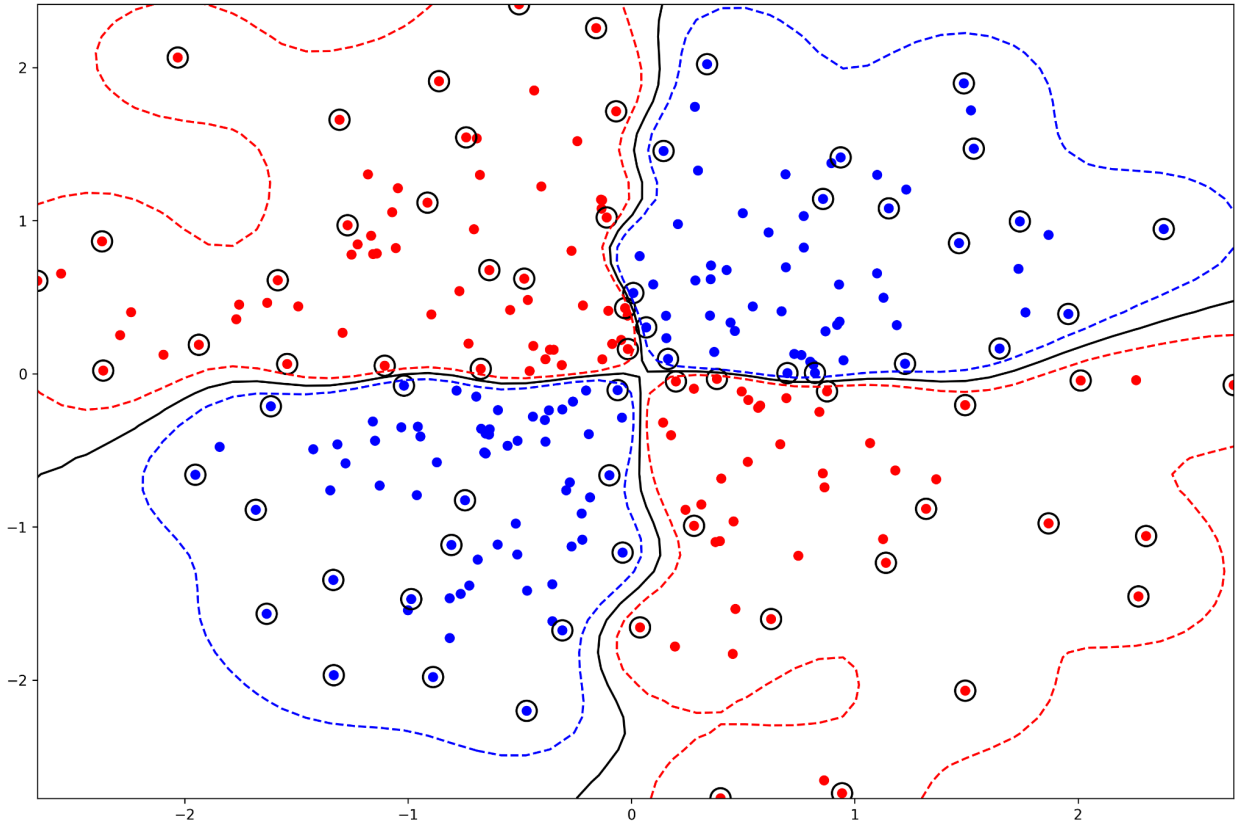
Dual SVM with polynomial kernel degree 6:



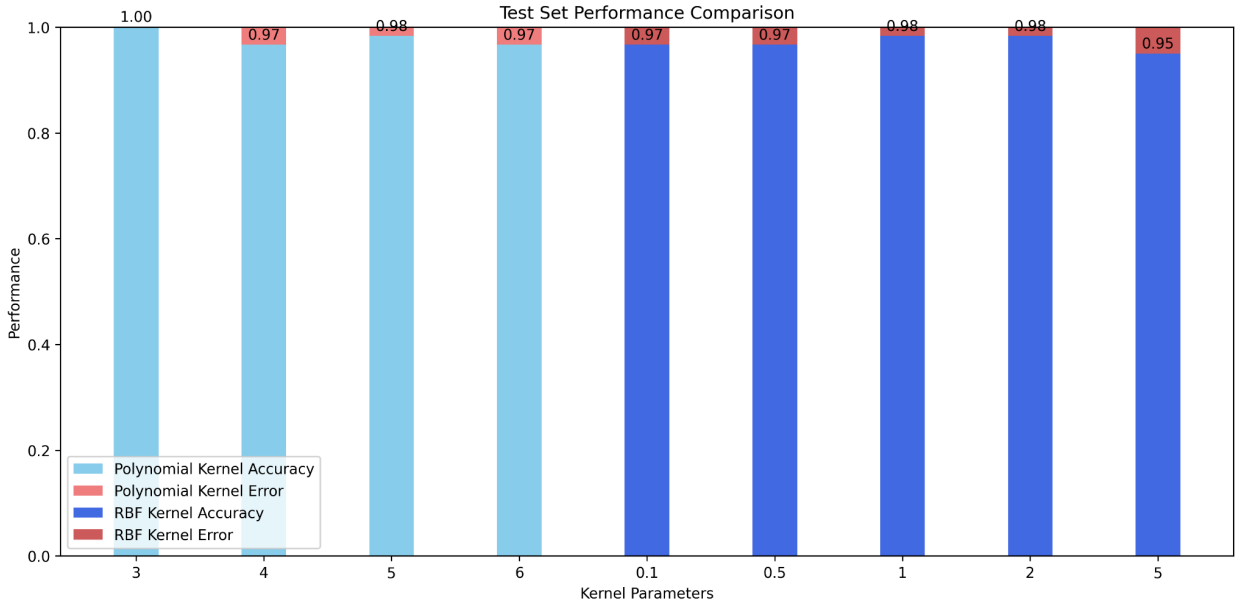
Dual SVM with RBF kernel gamma 2:



Dual SVM with RBF kernel gamma 5:



Error plot per model:



In this bar plot we can see the models performance and error comparison for the different kernels with different parameters(degree for polynomial and gamma for RBF).

The best model seem to be polynomial with degree 3.
This Bar plot demonstrates the comparison better than the ROC curve, I also create the ROC plot and it's available in the zip and its creation in the code.

Question 3

Output

No output - solution in the code

Output plots:

No output plots

Explanation

The code is an implementation of SVM class containing init, fit, predict, decision_function and score.
The class enable the kernels types: 'polynomial', 'rbf' or 'sigmoid'.

Question 4

Output

Testing different SVM kernels and parameters:
==========

Testing polynomial kernel:

Degree: 2, Accuracy: 0.9211, AUC: 0.9840
Degree: 3, Accuracy: 0.9386, AUC: 0.9895
Degree: 4, Accuracy: 0.7719, AUC: 0.8732
Degree: 5, Accuracy: 0.6579, AUC: 0.7665

Testing rbf kernel:

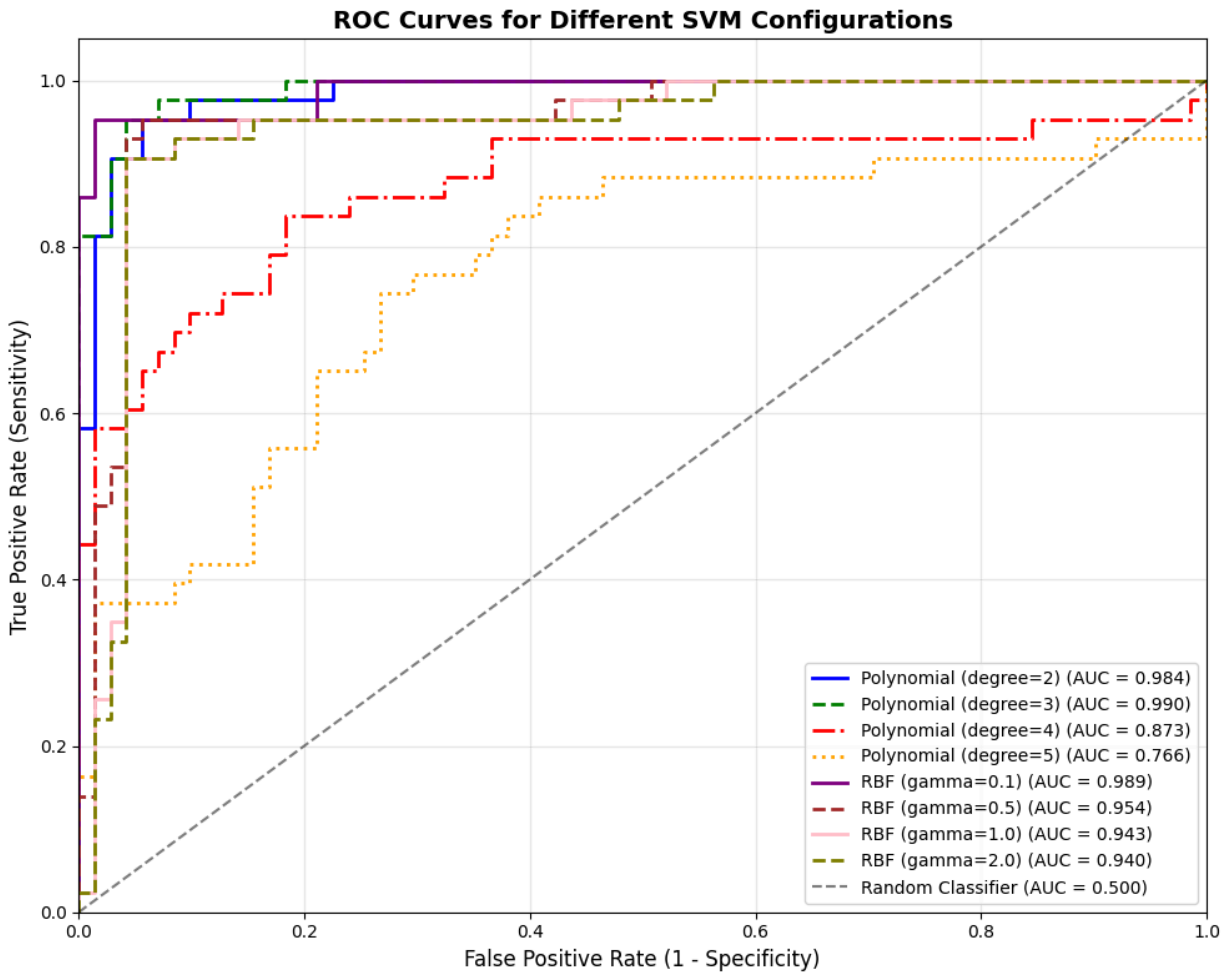
Gamma: 0.1, Accuracy: 0.9737, AUC: 0.9889
Gamma: 0.5, Accuracy: 0.9474, AUC: 0.9541
Gamma: 1.0, Accuracy: 0.9386, AUC: 0.9430

Gamma: 2.0, Accuracy: 0.9386, AUC: 0.9401

==========
SUMMARY OF ROC ANALYSIS:
==========
Polynomial (degree=2) | AUC: 0.9840
Polynomial (degree=3) | AUC: 0.9895
Polynomial (degree=4) | AUC: 0.8732
Polynomial (degree=5) | AUC: 0.7665
RBF (gamma=0.1) | AUC: 0.9889
RBF (gamma=0.5) | AUC: 0.9541
RBF (gamma=1.0) | AUC: 0.9430
RBF (gamma=2.0) | AUC: 0.9401

Best performing model: Polynomial (degree=3) (AUC = 0.9895)

Output plots:



Explanation:

Question 4 tests different SVM kernels and parameters.

The best performing model is Polynomial with degree 3 (AUC = 0.9895).

The analysis involved testing both polynomial and Radial Basis Function (RBF) kernels with different parameter values (degree for polynomial and gamma for RBF). The performance was measured using Accuracy and Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC).

The summary highlights the AUC values for each combination of kernel and parameter. Based on the AUC values, the Polynomial kernel with a degree of 3 showed the best performance, having the highest AUC score.