**Homework 4**

**Instructions**

This homework contains **6** concepts and **9** programming questions. In MS word or a similar text editor, write down the problem number and your answer for each problem. Combine all answers for concept questions in a single PDF file. Export/print the Jupyter notebook as a PDF file including the code you implemented and the outputs of the program. Make sure all plots and outputs are visible in the PDF.

Combine all answers into a single PDF named andrewID\_hw4.pdf and submit it to Gradescope before the due date. Refer to the syllabus for late homework policy. Please assign each question a page by using the “Assign Questions and Pages” feature in Gradescope.

Problem 1 (1 points)

Which of the following two models represents a better discriminator?

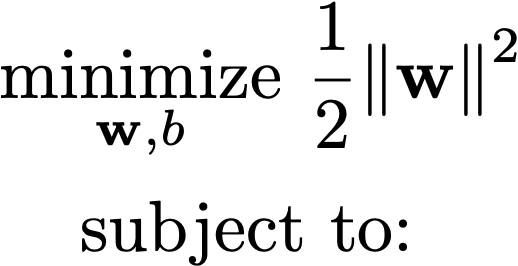
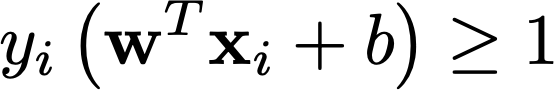
A diagram of a model

Description automatically generated

Problem 2 (1 points)

Multiple Choice (select one)

Consider an SVM classifier:



We would like to solve the problem with a quadratic programming solver:

A black background with a black square

Description automatically generated with medium confidence

When inputting the inequality constraint for quadratic programming packages, how should G and h be formulated? Consider

1. G = y\*[x1,x2,1], h = 1
2. G = -y\*[x1,x2,1], h = 1
3. G = y\*[x1,x2,1], h = -1
4. G = -y\*[x1,x2,1], h = -1

Problem 3 (2 points)

Which of the following two trained models will be faster to evaluate a set of 1000 test points?

A screenshot of a diagram

Description automatically generated

Problem 4 (2 points)

(Multiple Choice - select one)

Visually, which of the following SVM models classifies the data best?

1. Linear
2. Polynomial
3. RBF

A diagram of a person's body

Description automatically generated

Problem 5 (2 points)

Consider a multiclass SVM which classifies between 3 different classes. Each class has the same number of data points. Which would be faster to train, a one-versus-one or one-versus-rest classifier?

Problem 6 (2 points)

The following SVR model is fit to the data with an RBF kernel. Assume the model uses epsilon insensitive loss, Lϵ. How many data points contribute to the loss Lϵ for the given model?

