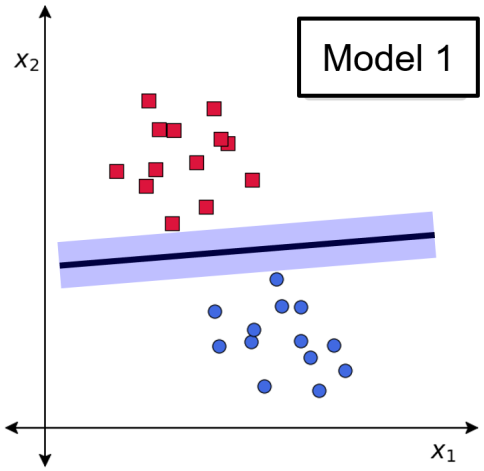
Homework 4

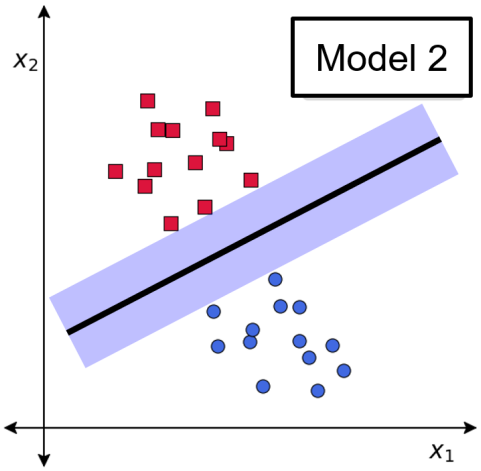
Instructions
This homework contains **6** concepts and **9** programming questions. In MS word or a similar text editor, write down the problem number and your answer for each problem. Combine all answers for concept questions in a single PDF file. Export/print the Jupyter notebook as a PDF file including 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.

Question	**Points**
Concept 1	2
Concept 2	2
Concept 3	2
Concept 4	2
Concept 5	2
Concept 6	2
M4-L1-P1	6
M4-L1-P2	6
M4-L1-P3	6
M4-L2-P1	6
M4-L2-P2	6
M4-L2-P3	6
M4-HW1	24
M4-HW2	24
M4-HW3	24
Total	**120**
Bonus	6

Problem 1
Which of the following two models represents a better discriminator?





Problem 2
Multiple Choice (select one)

Consider an SVM classifier:

$$\underset{\mathbf{w},b}{\text{minimize}}\ \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{subject to: } y_i\left(\mathbf{w}^T\mathbf{x}_i + b\right) \geq 1$$

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

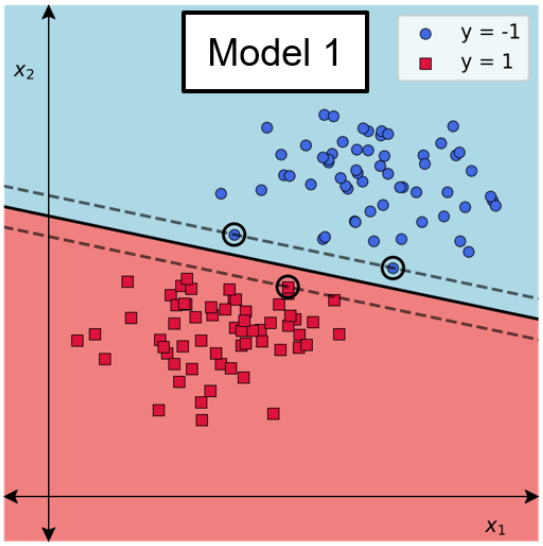
$$\begin{array}{ll}\text{minimize} & (1/2)x^TPx + q^Tx \\ \text{subject to} & Gx \preceq h \\ & Ax = b\end{array}$$

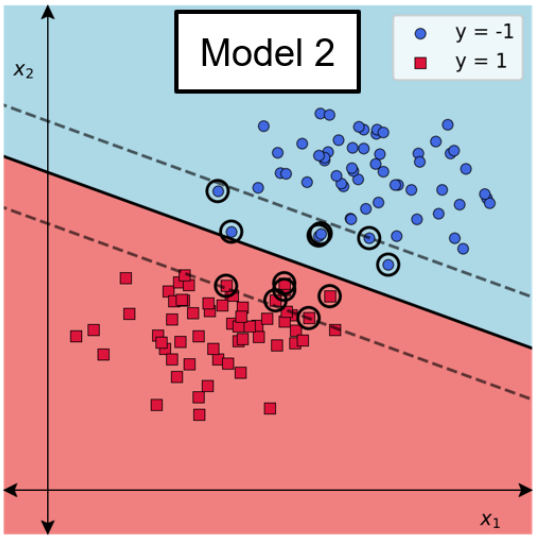
When inputting the inequality constraint for quadratic programming packages, how should G and h be formulated? Consider

1. G = y*[x_1,x_2,1], h = 1
2. G = -y*[x_1,x_2,1], h = 1
3. G = y*[x_1,x_2,1], h = -1
4. G = -y*[x_1,x_2,1], h = -1

Problem 3

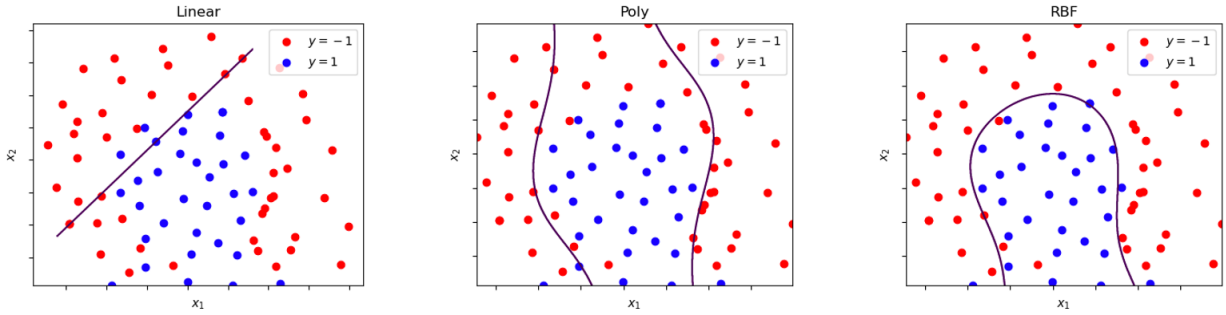
Which of the following two trained models will be faster to evaluate a set of 1000 test points?





Problem 4
(Multiple Choice - select one)
Visually, which of the following SVM models classifies the data best?

1. Linear
2. Polynomial
3. RBF



Problem 5

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

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