

**Instructions for homework submission**

a) There are two sections in this homework:

1. Write the solution to the first section in Latex.
2. For the programming questions, explain your thought process, results, and observations in a markdown cell after the code cells. Do not just include your code without justification. You can use available libraries for this question.

b) You need to submit a zip file:

- The name of the .zip file: FirstName\_LastName\_HW3.zip
- The zipped folder includes the following files:
  - A pdf (generated using Latex) for the math section: FirstName\_LastName\_HW3.pdf
  - A .ipynb jupyter notebook file for the programming section: FirstName\_LastName\_HW3.ipynb
  - A csv file with one main column that has the predictions with the name pred\_svm.  
File name: FirstName\_LastName\_preds.csv

c) Start early :)

d) Total: 100 points

**Math: Question 1 - 30 points**

We are given  $n = 7$  observations in  $p = 2$  dimensions. For each observation, there is an associated class label.

Index	$X_1$	$X_2$	Y
1	3	6	Blue
2	2	2	Blue
3	4	4	Blue
4	1	3	Blue
5	2	0	Red
6	4	2	Red
7	4	0	Red

- (a) Sketch the optimal separating hyperplane, and provide the equation for this hyperplane.
- (b) Describe the classification rule for the maximal margin classifier. It should be something along the lines of "classify to Red if  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 \geq 0$ , and classify to Blue otherwise." Provide the values for  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ .
- (c) On your sketch, indicate the margin for the maximal margin hyperplane.
- (d) Indicate the support vectors for the maximal margin classifier.
- (e) Does a slight movement of the seventh observation affect the maximal margin hyperplane? Justify your answer.
- (f) Draw an alternative hyperplane that is not the optimal separating hyperplane, and provide the equation for this hyperplane.

(g) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

### Math: Question 2 - 20 points

Assume that we have a training data with four samples, 2 features and 2 classes. The positive examples are (1,1) and (-1,-1). The negative examples are (1,-1) and (-1,1).

(a) Draw a table that represents this training set. What is the shape of  $X$  and  $y$ ?

(a.bonus) The table you just drew is called a truth table. Do you know the logic gate representing this truth table?

(b) Draw these four points on x-y plane. Are these points linearly separable?

(c) Consider the feature transformation  $\phi(x) = [x_1, x_2, x_1x_2]$ , where  $x_1$  and  $x_2$  are, respectively, the first and second coordinates of a generic example  $x$ . Draw these four points when transformed by the function  $\phi(x)$ . Are these four transformed points now linearly separable?

(d) What is the margin size after the transformation? Which points are support vectors?

### Programming: Question 3 - 50 points

In this problem, we will use university application data for the purpose of admission classification. Fine `data_train` and `data_test` on canvas.

(a) **Data Pre-processing:** Create a binary label based on the column “*Chance of Admit*”. Convert any values bigger than the median to 1 and 0 otherwise. Split the data into training and validation sets. You can use a 80-20 split.

(b) **Model Initialization:** Initialize 3 different SVM models with the following kernels.

1. Linear kernel
2. RBF kernel
3. Polynomial (degree 3) kernel

(c) **Feature Selection and Model Training:** Train each SVM Model above with the following feature combinations to predict admission.

1. CGPA and SOP
2. CGPA and GRE Score
3. SOP and LOR
4. LOR and GRE Score

(d) **Support Vectors:** What are the support vectors for each model and feature combination? How many support vectors does each class have in each case?

(e) **Result Visualization:** For each kernel - input combination, visualize the predictions on the training set. Color code the points by their labels.

(f) **Result Analysis:** Just by looking at the figures you generated, answer this question: Which features + kernel combinations give you the best result? Validate your model on the validation set and find the best performing combination with respect to accuracy.

(g) **Inference:** Use the best model you found in the previous step to predict the label of the test data. Save the prediction in a csv file “FirstName LastName preds.csv”