

# Assignment 7

## CS 750/850 Machine Learning

- **Due:** April 7th at 11:59PM
- **Submission:** Turn in both a **PDF** and the **source code** on MyCourses
- **Questions:** Piazza
- **Office hours:**
  - Marek: Wed 1:30-2:30pm <https://unh.zoom.us/j/356262933>
  - Soheil: Mon 2pm-4pm
  - Xihong: Thu 1:30-3:30pm <https://unh.zoom.us/j/587656408>

### Problem 1 [33%]

Here we explore the maximal margin classifier on a toy data set.

- (a) We are given  $n = 7$  observations in  $p = 2$  dimensions. For each observation, there is an associated class label. Sketch (*by hand is OK*) the observations.

Obs.	$X_1$	$X_2$	$Y$
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

- (b) Sketch (*by hand is OK*) the optimal separating hyperplane, and provide the equation for this hyperplane (of the form (9.1)).
- (c) Describe the classification rule for the maximal margin classifier. It should be something along the lines of “Classify to Red if  $\beta_0 + \beta_1 X_1 + \beta_2 X_2 > 0$ , and classify to Blue otherwise.” Provide the values for  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ .
- (d) On your sketch, indicate the margin for the maximal margin hyperplane.
- (e) Indicate the support vectors for the maximal margin classifier. How will the number of support vectors depend on the dimensionality of the space.
- (f) Argue that a slight movement of the seventh observation would not affect the maximal margin hyperplane.
- (g) Sketch a hyperplane that is not the optimal separating hyper-plane, and provide the equation for this hyperplane.
- (h) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane.

## Problem 2 [33%]

Generate a simulated two-class data set with 200 observations and two features in which there is a visible but non-linear separation between the two classes. Explore whether in this setting, a support vector machine with a polynomial kernel (with degree greater than 1) or a radial kernel will outperform a support vector classifier on the training data. Which technique performs best on the test data? Make plots and report training and test error rates in order to back up your assertions.

## Problem 3 [33%]

Apply SVMs and at least 3 different kernels to a data set of your choice. Use cross-validation to optimize the parameter  $C$ . Be sure to fit the models on a training set and to evaluate their performance on a test set. How accurate are the results compared to simple methods like linear or logistic regression? Which of these approaches yields the best performance?