HW #2 Due: 4/10/2020

1. Implement the k-NN classifier for Iris dataset. To begin one experiment, randomly draw 70 % of the instances for training and the rest for testing. Repeat the drawing and the k-NN classification 10 times and compute the average accuracy. Then, plot the curve of k versus accuracy for k = 1, 3, …, 15. For simplicity, use the Euclidean distance in your computation.
2. Follow problem 1, but use Bayes optimal classifier instead. To do so, you need to compute the sample mean and sample covariance for each of the class. When you are done, compare the accuracy between k-NN, Naïve Bayes (in HW #1), and Bayes optimal classifier.
3. Consult the Internet to learn how to perform regression based on k-NN, and use the first three features (sepal length, sepal width, petal length) in each sample as input to predict the fourth feature (petal width) in the Iris dataset. To conduct one trial, again you need to divide the dataset into a training set (70%) and a test set (30%). Remember, you need to build one regression model for each class. Repeat 10 trials and plot the curve of k versus average MSE for each class for k = 1, 3, …, 15. Is k-NN regression better than linear regression (in HW #1)?
4. Use Naïve histogram estimator to plot $\hat{p}(x)$ for feature "sepal length" in the Iris dataset. Use all three classes in the plot with the number of bins $N = 100$. Do you observe three peaks for three classes in the plot? Hint: (a) To determine the bin width, you need to know the range of values in the dataset. (b) To make $\hat{p}(x)$ a smooth curve, you need at least 10 $\hat{p}(x)$ values for x within a bin.
5. In the lecture, we mentioned that $\sum_{j=1}^{J}|\beta_j| \le c_0$ is a regularization term in Lasso.
   i. Use your own words to explain why this term "regularizing" β_j (i.e., so that β_j don't take extreme values).
   ii. If we implement two regression programs, one using $\sum_{j=1}^{J}\left|\beta_j\right| \le 1$ and the other one using $\lim_{n\to\infty}\sum_{j=1}^{J}{\beta_j}^n \le 1$ as the regularization term, which one do you expect to have lower bias? Which one has lower variance? Hint: Simple models are more likely to have high bias, whereas complicated models are more likely to have high variance.