ACTU PS5841 Data Science in Finance & Insurance - Spring 2022 (Y. Wang)
Assignment - 7.5
Assigned 3/10/22, Due 3/20/22 (Sun)

**Problem 1.** KNN
$X_1 + \sin X_2 - 5 = 0$, where $X_1 \in [0, 10]$ and $X_2 \in [0, 10]$ defines a boundary in the feature space. In addition, the true class probabilities are

$$\Pr(Y = 1|\boldsymbol{x} \text{ is on the good side of the boundary}) = 0.8 + (x_1 + x_2) * 0.01$$

$$\Pr(Y = 2|\boldsymbol{x} \text{ is not on the good side of the boundary}) = 0.8 + (x_1 + x_2) * 0.01$$

Feel free to define which side of the boundary is the good side.

[1] Produce a chart which shows, on the feature space $\{(X_1, X_2)\}$, the decision regions of the Bayes Classifier.

[2] Code to estimate the overall Bayes error rate

[3] Generate a sample, consitent with the true class probabilities, on the grid where $X_1 \in \{0, 0.2, 0.4, 0.6, ..., 10\}$ and $X_2 \in \{0, 0.2, 0.4, 0.6, ..., 10\}$.
Produce a chart which shows, on the feature space $\{(X_1, X_2)\}$, the decision regions of a KNN Classifier with $K = 5$.

[4] Code to estimate the test error rate of the KNN Classifier with $K = 5$, using the 5-fold cross validation approach. Report the following

| Validation Set | Estimated Test Error Rate for this Validation Set |
|---|---|
| 1 | |
| 2 | |
| 3 | |
| 4 | |
| 5 | |
| Estimated KNN(5) Error Rate | |

You may find the following resources useful.
`https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html` `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html`