

Texas Christian University
COSC 40023 Spring 2025
Assignment 7

Due: April 23rd, Wednesday, at 11:30 PM (Late submission NOT accepted)

Submission (two files, no compression): `assignment7.py` and `assignment7.R` through TCU Online

Download the dataset called “Iris_Data.csv”. The dataset contains three classes of 50 samples each, where each class refers to a type of iris plant. Two features are measured from each sample: the length and the width of the sepals, in centimeters.

Build SVM models in both Python and R. Two files named `assignment7.py` and `assignment7.R` should be created. Here are some additional requirements.

1. In the preprocessing part, no need to split the dataset into a training set and a test set. Use the whole dataset to train the model.
2. Feature scaling is required for both Python and R.
3. You must try four kernels and choose the one that performs the best based on their accuracy. The four kernels are linear, polynomial with a degree of 3, radial basis and sigmoid. After you found the best performance kernel, ONLY keep the code for this kernel. In addition, keep the code for printing the confusion matrix and accuracy of the best performance kernel.
4. A plot should be generated in each programming language. In a plot:
 - The horizontal axis should be scaled sepal length and the vertical axis should be scaled sepal width.
 - For background regions, use three different colors of your choice to represent the three classes.
 - For observation dots, use different transparencies or shades of the above three colors to represent the three classes. For example, under the same class, if the background is red, the dots should be a different transparency or shade of red.
 - Have proper title and axis labels.
5. Have sufficient single-line and multi-line comments.