**CSc 44700, Spring 2024**

**Assignment 3**

**Due March 28, 2024, at midnight**

This assignment involves the Iris Dataset. If you want, you can use portions of the sample code that I’ve posted to load the data dataset and draw some plots, but you don’t have to.

For the set of all of the categories (Setosa, Versicolor, Virginica) of the data listed above, create a series of models (listed below) for the combination of the two features “Sepal Length” and “Sepal Width ”

Separate the data (the 150 data samples) into a training set and a testing set, using the Scikit learn function “train\_test\_split,”

XX\_train, XX\_test, YY\_train, YY\_test

= train\_test\_split(XX, YY, test\_size=0.20, random\_state=XYZ),

where “XYZ” are the last three digits of your student id.

Apply the following models to the training and testing data:

1. Stochastic gradient descent using the hinge loss function,
2. Stochastic gradient descent using the modified huber loss function,
3. Logistic regression using the parameters of your choice,
4. Support vector machine using the linear kernel,
5. Support vector machine using the rbf kernel,
6. Decision tree using the gini impurity as the impurity measure,
7. Random forest using the parameters of your choice, and
8. KNN using two values of *k* of your choice,

For each of these models calculate the accuracies of your model with the training data and with the test data.

Present your findings in the clearest way that you can; including relevant plots is very strongly encouraged.

**Discuss your results.** Include in your discussion how your choice of model affects the accuracies of the model on the training data and on the testing data.

Please submit your work on Blackboard as an ipynb file. Please name your file as

LastName\_FirstName\_AS03.ipynb

You must discuss your answers and describe how you came up with them. Show your work. Just stating an answer won’t get you more than half credit.

If you collaborate on this or any other assignment, you must have contributed substantially to anything you submit; just using a current (or past) classmate’s work without having contributed substantially to it is not collaboration -- it’s cheating.

If you collaborate with anyone you must indicate with whom you collaborated.