

For this in-class assignment, we will use the very popular iris dataset. It includes three iris species with 50 samples each as well as some properties about each flower. One flower species is linearly separable from the other two, but the other two are not linearly separable from each other. Consider the `Iris.csv` data file posted on Blackboard (under the In-Class 5 assignment link). **In Python**, answer the following:

1. (3 points) Using the pandas library, read the csv data file and create a data-frame called `iris`.
2. (4 points) Using the `where` function from `numpy`, create a new variable called `Species_numb` as follows:
  - if `Species = Iris-virginica`, then `Species_numb = 1`
  - if `Species = Iris-versicolor`, then `Species_numb = 2`
  - if `Species = Iris-setosa`, then `Species_numb = 3`
3. (5 points) Using `SepalLengthCm`, `SepalWidthCm`, `PetalLengthCm`, and `PetalWidthCm` as the input variables, and `Species_numb` is the target variable, split the data into two data-frames (taking into account the proportion of 1s, 2s and 3s): train (80%) and test (20%).
4. (3 points) Standardize the input variables in the train and test datasets. Use the 0-1 scale.
5. (8 points) Using train dataset and the one-vs-one multi-class classification strategy with the logistic regression model, build a multi-class classification model. Then, use this model to make predictions on the test data. Compare these predictions with the actual values using the confusion matrix.
6. (8 points) Using train dataset and the one-vs-one multi-class classification strategy with the decision tree model (with maximum depth equal to 3), build a multi-class classification model. Then, use this model to make predictions on the test data. Compare these predictions with the actual values using the confusion matrix.
7. (3 points) Using the results from part 5 and 6, what model would use to predict iris species? Be specific.