

**CSc 447000, Fall 2022**  
**Assignment 2**  
**Due September 26, 2022**

This assignment involves the Iris Dataset. I recommend, but don't require, that you use the sample code that I've posted to load the data dataset and draw some plots.

As we discussed in class, the dataset includes 150 samples, 50 from each of three species,

- $\alpha$ ) Setosa
- $\beta$ ) Versicolor
- $\gamma$ ) Virginica

For each of the samples, the dataset includes four features,

- A) sepal length
- B) sepal width
- C) petal length
- D) petal width

For a classification problem using the species as the classes, we can load the data for two or three of the classes and one, two, three, or four of the features.

Two classes

1.  $\alpha\beta$
2.  $\alpha\gamma$
3.  $\beta\gamma$

Three classes

1.  $\alpha\beta\gamma$

We can choose to use, one, two, three, or all four features in a model.

One feature

1. A
2. B
3. C
4. D

Two features

1. AB
2. AC
3. AD
4. BC
5. BD
6. CD

Three features

1. ABC

2. ABD
3. ACD
4. BCD

Four features

1. ABCD

For the subset  $\alpha\beta\gamma$  (Setosa, Versicolor, Virginica) of the data listed above, create a model for each of the 15 possible combinations of features. For each of these 15 combinations of features, determine the accuracy of the model using the Scikit Learn library and its SGD classifier using four different loss functions (any four that you want to use).

Present your findings in the clearest way that you can, comparing the accuracy of 60 features/loss-function pairs. Including relevant plots is encouraged.

Discuss your results. Include in your discussion how varying the combinations of features used and the loss function chosen affect the accuracy of the model.

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

LastName\_FirstName\_AS02.ipynb

You must discuss your answers and describe how you came up with them. Show your work. Just stating a correct 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.