



Iris Classification Project

Rajas Gupta | rajasg@umich.edu **Vijay Sharma** | vscharm@umich.edu

OVERVIEW

During this project we will train a logistic regression classifier to predict species of iris based on sepal length, sepal width, petal length, and petal width. We will then tweak certain aspects of our model to see how the accuracy of the resulting classifier is affected.

STEPS

1. Follow the tutorial found [here](#)

- Try to understand what the code is doing and make comments for yourself
- You may get a `ConvergenceWarning` upon running the program. This can be ignored
- The very last number printed (accuracy) should be 1.0

2. Find pair a features to train on that result in accuracy of less than 1.0

- You can estimate the predictive power of each feature by inspecting the scatterplot. For instance, look at the red for versicolor and setosa. You should be able to estimate how useful `petal_length` would be for differentiating between these two species.
- **Hint:** `X = data.iloc[:, [2, 3]]` selects the features `petal_length` and `petal_width`
- Remember you will have to re-run your code to fit your classifier on your new feature set

3. Change proportion of test points and type of classifier

- Reset `X = data.iloc[:, :-1]` so it again contains all the features
- Set `test_size=0.5` in the `train_test_split()` function call
- Re-run your code. Does accuracy change?
- Run `from sklearn.linear_model import SGDClassifier`
- Set `model = SGDClassifier()`
- Re-run your code. Does accuracy change?

4. Upload your code [here](#)
