

CS173 Class Activity – Apr 29, 2019

Classification

1. Download and install Jupyter. I suggest trying the Anaconda version, but you may do as you will as long as it works. I am not particular (i.e., don't care) what version of Python you prefer, 2.7 or 3. Modify your work accordingly. I use Python 2.7, fyi.
<https://jupyter.org/install>
2. Learn Jupyter on your own. Try the tutorial. Try a few more tutorials. Learn about code versus markdown. Don't be lazy about it.
3. Read about and learn about the famous Iris dataset. Again, don't be lazy about it.
4. Download the iris dataset. I might suggest the CSV version I use on the Share Drive if you want to be consistent with my sample code below. But you can modify the code accordingly if you prefer other versions of the dataset.
<http://bit.ly/ucrCS173sp19share>
5. Build a classifier for the Iris dataset. Use the sample Jupyter project in the Share Drive, it's called "Iris Classifier.ipynb" btw.
6. Follow the code carefully, step-by-step. Think while you work. Think more. Don't be lazy about it. Try, fiddle, experiment.
7. You should have learned from this exercise a few things, if you didn't know them already:
 - a. Pandas and dataframes
 - b. how to build a regression and random forest classifier in python
 - c. how to fiddle with features and examine their effects
 - d. how to evaluate classifier performance
 - e. how AUC works and what it means
8. Once you're satisfied you learned enough, ask the TA to give you credit for completing the activity and enter your score. Then upload to gradescope. Groups up to 3 is okay.

SCORE: 10

There will be a QUIZ to test what you learned.