

# Chapter 6 Exercises

*Erin M. Buchanan*

*03/12/2019*

## Get Started

- Create a Jupyter notebook with the following items. You should upload a compiled version of the notebook in html or pdf format.
- Remember, use Markdown cells to answer text questions. Paste the questions into the cells so it's clear what you are answering.

## Classify Names

- We saw in our first lecture that we could classify the gender of names based on the last letter or two of a person's name. Using any *other* features you can think of, build the best name gender classifier you can.
- Split the names corpus into three subsets: 6900 for training, 500 for dev-test, and 500 for test.
- Create a classifier that includes more than just the last letter of the person's name (i.e., it should have more features than the examples performed in class).
- Use the dev-test to examine your incremental progress.
- Once you are satisfied, test it against the test dataset.
- Explain it's performance:
  - Describe what you used as your feature set.
  - How well did it perform? Include the accuracy of your classifier.
  - What were the most important features? Include the most informative features and discuss.
  - Include a confusion matrix. What can you tell about your errors and correct guesses from this matrix?
  - Calculate precision, recall, and the F-score for your classifier for both guesses.