

UCLA Extension Data Science Intensive

Instructor: William Yu

Project 6

A. Python on Machine Learning Models

- In Project 4 (see P04.R), we analyzed the breast cancer data by testing various machine learning models.
- Now use Python Spyder to do a similar work (including missing value imputation and 10-fold cross validation) and show the best machine learning model.
- *Note: use py01.py, py03.py and Google.*

B. Python and Jupyter: What Factors Predict Movie Blockbusters?

- Open the py06_movie.py Python script and import the data movie_test.csv and movie_train.csv.
- Go through the script and understand the Python script. At the end, you will run a simple linear regression for the train set. Repeat the same procedure in Jupyter and explain the results and save it as a html file (click File → Download As → HTML). Submit the Jupyter html file.