CS 210 RECITATION CONTENTS:

Below is a quick reference for the topics demonstrated using Python during recitations throughout the semester. Chances are you have already seen code similar to what you need for your project. If not, a basic Google search like "*train test split SKLearn*" will provide code which you can use as a template.

Recitation 1

- Python Basics:
  - Lists, tuples, dictionaries
- Graphing Basics:
  - Matplotlib's plot function

Recitation 2

- More Python:
  - Regular expressions
  - Command line operations
  - Reading files
  - Using the Twitter, GitHub APIs
  - Scraping the web

Recitation 3

- Introduction to NumPy Library:
  - Conversions among Numpy ndarray other array-like objects
  - Filtering array elements
  - Matrix operations
- Introduction to Pandas Library:
  - Selecting dataframe rows by specifying criteria on column values
  - 1-D, 2-D histograms
  - Scatter plots

Recitation 4

- Introduction to SKLearn Machine Learning Library:
  - K-Nearest Neighbors (KNN) Classifier

Recitation 5 & 6

- Students started working on projects during recitation hours, there were no demos

Recitation 7

- Based on feedback from recitations 5 & 6, detailed data handling with Pandas:
  - Selecting dataframe contents based on both columns and rows
  - Re-indexing rows
  - Renaming, merging, transforming, dropping columns
  - Changing the content and the type of fields
  - Transforming or dropping null values
  - Sorting, grouping data
  - Getting basic dataframe statistics like counts, min-max values
  - Reporting and reducing memory usage

Recitation 8

- Calculating and visualizing correlations
- Creating heatmaps using Seaborn Library
- Creating interactive plots with Bokeh
- Biclustering with spectral co-clustering

Recitation 9

- Creating pair plots and regression plots using Seaborn Library
- Calculating and interpreting rank correlations:
  - Spearman's Rho
  - Kendall's Tau
- Testing statistical significance using p-values with the help of SciPy Library
- Interactive visualization of rank correlations using parallel axis:
  - D3JS demo

Recitation 10

- Naive Bayes classification using SKLearn
- Checking accuracy by calculating the confusion matrix
- Real-world document classification example:
  - Feature extraction using bag-of-words model
  - Training a classifier
  - Evaluating accuracy by cross-validation
  - Categorization of 55000 e-mails into SPAM or HAM

Recitation 11

- Manual calculation of Information Gain to select attributes to make effective splits
- Decision Tree classification and regression using SKLearn
- Visualizing the resulting decision tree rules using GraphViz
- Visualizing the decision tree surface

Regression Lecture Demo

- Estimating regression coefficients using Statsmodels Library
- Plotting the least-squares line
- Reporting confidence intervals
- Testing the hypothesis of the existence of relationships between input-output variables with the help of p-values
- Reporting goodness of fit by calculating R-squared
- Feature selection and multiple linear regression
- Handling categorical predictors