

Homework Assignment 1

Due: 11:59pm, Feb 16, 2018

The goal of this problem set is to become familiar with basic Python commands, practice manipulating vectors and matrices, and try out basic data processing and plotting functions. If you are new to this language, please read the attached python reference book as well as the slides.

Question 1

In python, the library `matplotlib.pyplot` is a **collection** of command style functions that make `matplotlib` work like MATLAB. Each `pyplot` function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc. In `matplotlib.pyplot` **various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area**, and the plotting functions are directed to the current axes (please note that “axes” here and in most places in the documentation refers to the *axes* part of a figure and not the strict mathematical term for more than one axis).

Sample Codes:

```
import numpy as np
import matplotlib.pyplot as plt

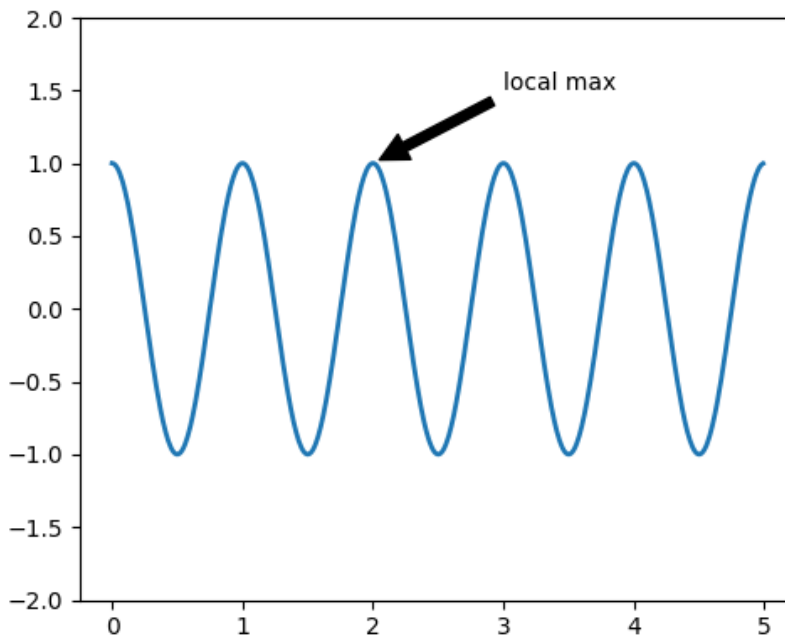
ax = plt.subplot(111)

t = np.arange(0.0, 5.0, 0.01)
s = np.cos(2*np.pi*t)
line, = plt.plot(t, s, lw=2)

plt.annotate('local max', xy=(2, 1), xytext=(3, 1.5),
            arrowprops=dict(facecolor='black', shrink=0.05),
            )

plt.ylim(-2,2)
plt.show()
```

Outputs:



Please select a proper range for the variable x , and visualize the curves for the following functions:

- straight line $y = \theta_0 + \theta_1 x$ where $\theta_0 = 30$, $\theta_1 = 0.5$
- quadratic function: $y = (x - \theta_1)^2 + \theta_0$, where $\theta_1 = 25$, $\theta_0 = 20$
- log function, $y = -\log(x)$ and $y = -\log(1 - x)$,
- sigmoid function, $y = 1/(1 + e^{-x})$

Question 2

2.1 (1 Credit) You are given the following tasks, all of which can be solved with a certain type of machine learning algorithms.

- Classify** a house to be single family or townhouse. **A training set is available**. Each training sample is provided with size, number of bed rooms, number of bathrooms, and house type (single family or townhouse).
- Classify an email to be spam or not. Users already identified some emails to be spam and labeled it for training.
- Human tumor Microarray data are provided as a matrix where row correspond to genes and columns to tissue samples. The task is to cluster columns (or samples) to identify disease profiles, i.e. tissues with similar disease should yield similar expression profiles.

Please select one of the following statements that are all correct:

- i) unsupervised learning with discrete predictions; ii) supervised learning with continuous predictions; iii) supervised learning with continuous predictions;
- i) supervised learning with discrete predictions; ii) supervised learning with discrete predictions; iii) unsupervised learning with discrete results;

discrete
refers to
output
i.e. length, not prob
- continuous

- c. i) supervised learning with discrete predictions; ii) supervised learning with continuous predictions; iii) unsupervised learning with discrete results;
- d. All three scenarios can be solved by unsupervised learning.

Question 3

To classify a fish to be a salmon or sea-bass, we collect a set of training samples, and measure the length of each sample fish. The category label of each training sample is also provided.

Which descriptions are correct? (Mark all when applicable)

- a. This problem can be solved by supervised learning that uses a single feature as predictor;
- b. This problem can be solved by unsupervised learning that uses a single feature as predictor;
- c. To solve this problem, we should select a complex model that can well classify the training samples;
- d. To solve this problem, we should compare different models, being complex or not, over both training data and testing data before making a selection.

Question 4 Please figure out and describe one machine learning task, either classification or regression, based on your own experiences and background. It should not be demo tasks we taught in class, e.g., hand digit recognition, house pricing, fish classification etc. Please design your machine learning algorithm for solving it, and answer the following questions. Note that you only need to submit your plans. No source codes or results required.

Task. Please explain briefly the inputs, outputs, and your goal in general.

Data preparation. Please describe how to collect dataset, including training data, validation data, and testing data. Please explain how to get the ground-truth label for all samples;

Submission instructions: what to hand in

- Prepare Single PDF file to address the above questions.
- ✓ Question 1: results and codes
- ✓ Question 2: choices
- ✓ Question 3: choices
- ✓ Questions 4: writings
- Submit your PDF file and source codes through the SDSU Blackboard system.
- No HARD COPY IS REQUIRED.