CS 596 Machine Learning

**Homework Assignment 1**

Due: 11:59pm, September 16, 2019

The goal of this problem set is to help you 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 books as well as the slides.

**Question 1**

In python, the library [matplotlib.pyplot](https://matplotlib.org/api/pyplot_api.html#module-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](https://matplotlib.org/api/pyplot_api.html" \l "module-matplotlib.pyplot" \o "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](http://matplotlib.org/faq/usage_faq.html#parts-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),

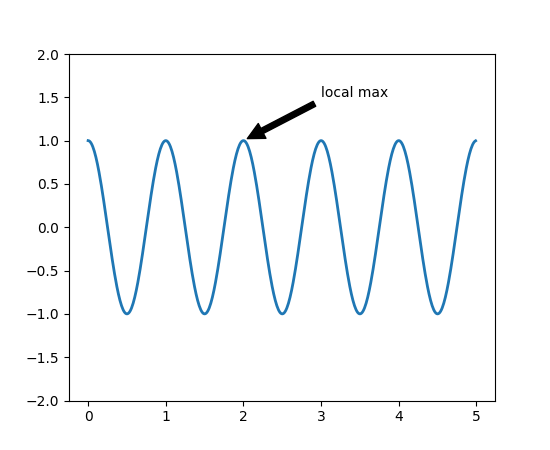
arrowprops=dict(facecolor='black', shrink=0.05),

)

plt.ylim(-2,2)

plt.show()

Outputs:



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

* Straight line where
* Quadratic function:
* Log function, and ,
* Sigmoid function,

**Question 2**

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

1. Classify a house to be single family or townhouse. A training set is available. Each training sample is provided with multiple features, including number of bed rooms, number of bathrooms, and house type (single family or townhouse).
2. Classify an email to be spam or not. Users already identified some emails as spam ones.
3. Human tumor Microarray data are provided as a matrix where rows correspond to genes and columns to tissue samples. The task is to cluster columns (or samples) to identify disease profiles: tissues with similar disease should yield similar expression profiles.

Which statement is correct? (Mark one)

1. i) unsupervised learning with discrete predictions;

ii) supervised learning with continuous predictions;

iii) supervised learning with continuous predictions;

1. i) supervised learning with discrete predictions;

ii) supervised learning with discrete predictions;

iii) unsupervised learning with discrete results;

1. i) supervised learning with discrete predictions;

ii) supervised learning with continuous predictions;

iii) unsupervised learning with discrete results;

1. All the three scenarios can be solved by unsupervised learning.

**Question 3** Please figure out and describe one machine learning task, either classification or regression, based on your own experiences and background. It should not be the demo tasks taught in class, e.g., hand digit recognition, house pricing, fish classification etc. Please design your machine learning algorithm, and answer the following questions. Note that you only need to submit your plans. No source codes or results are required.

*Task*. Please briefly explain 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: writings
* Submit your PDF file and source codes through the SDSU Blackboard system.
* No HARD COPY IS REQURIED.