CS596_Assignment_1

September 16, 2019

1 Question 1: Data Visualization (Results and Code)

[1]:
```
import numpy as np
import matplotlib.pyplot as plt
```

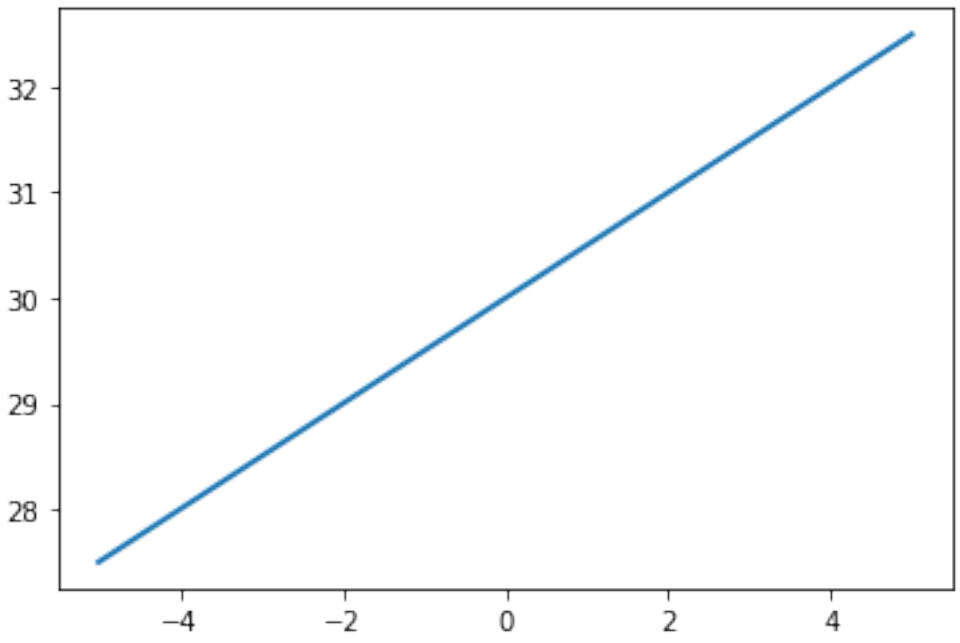
[2]:
```
# Straight Line

x = np.arange(-5.0, 5.0, 0.01)
t0 = 30
t1 = 0.5

y = t0 + (t1 * x)

line = plt.plot(x, y, lw=2)

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



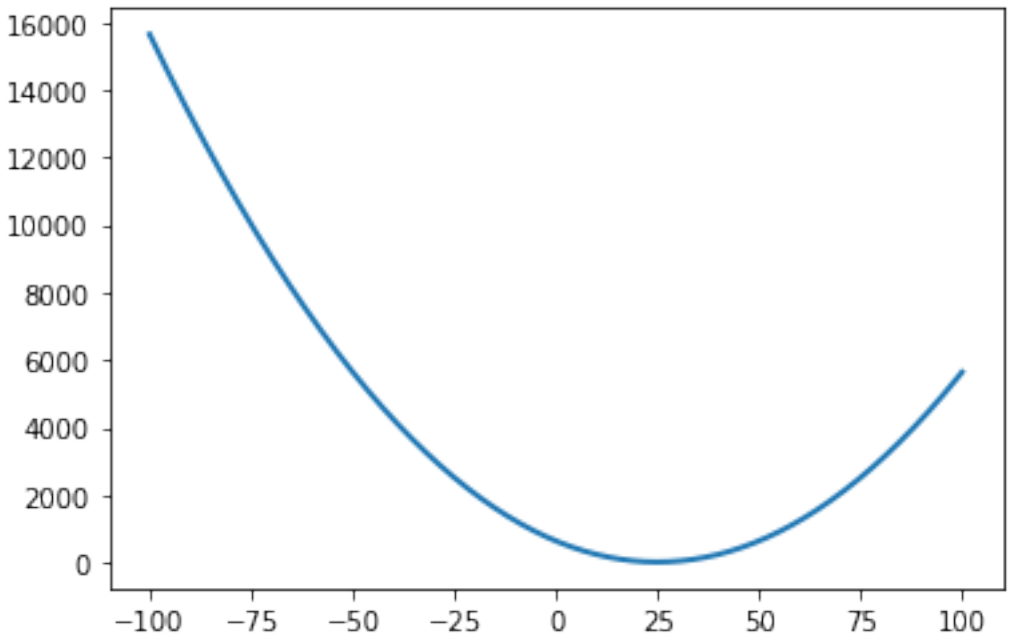
[3]:
```
# Quadratic Equation

x = np.arange(-100.0, 100.0, 0.01)
t1 = 25
t0 = 20

y = np.square(x - t1) + t0

line = plt.plot(x, y, lw=2)

plt.show()
```



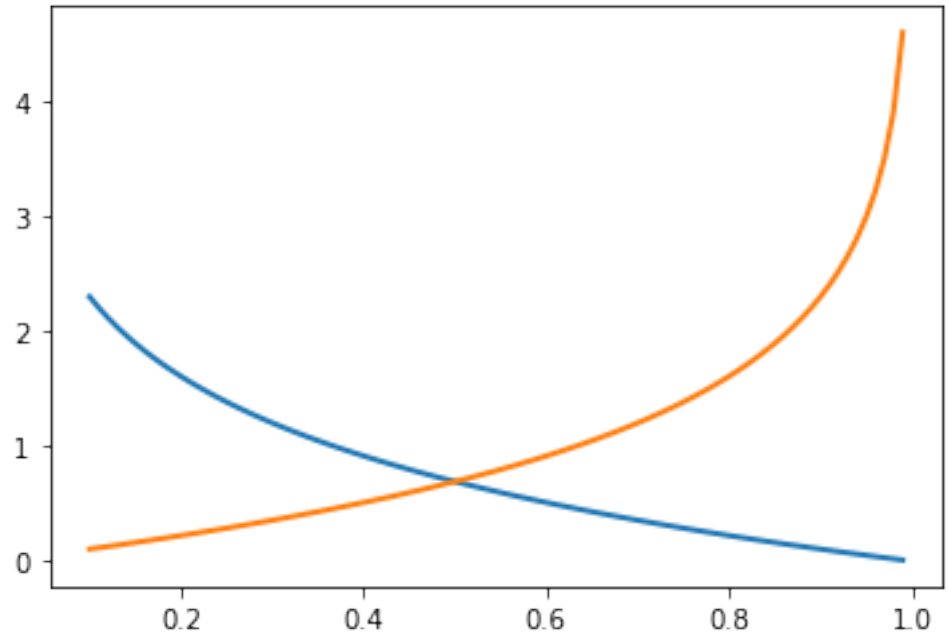
[4]:
```
# Log Function

x = np.arange(0.1, 1, 0.01)

y1 = -np.log(x)
y2 = -np.log(1-x)

line1 = plt.plot(x, y1, lw=2)
line2 = plt.plot(x, y2, lw=2)

plt.show()
```



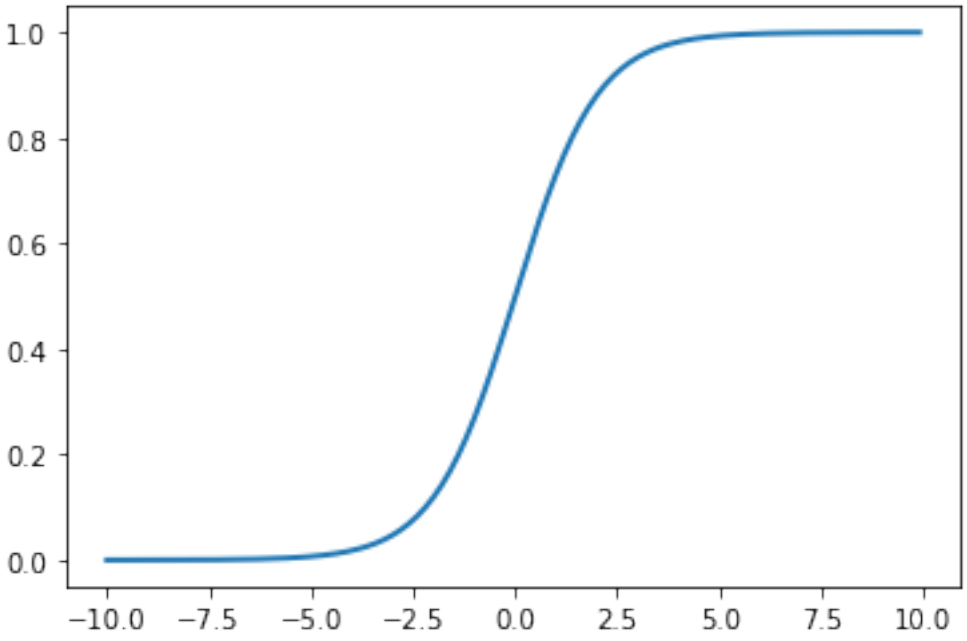
[5]:
```
# Sigmoid Function

x = np.arange(-10.0, 10.0, 0.1)

y = 1 / (1 + np.exp(-x))

line = plt.plot(x, y, lw=2)

plt.show()
```



2 Question 2: Machine Learning MCQs

The correct statement between the given options would be: **b**. `i) supervised learning with discrete predictions; ii) supervised learning with discrete predictions; iii) unsupervised learning with discrete results;`

3 Question 3: Describe one Machine Learning Algorithm

The machine learning algorithm chosen is a Regression based algorithm. I intend to limit the gun violence incidences caused in the US. The primary reason would be that being an international student from India, there are very less cases of gun violence or mass public shooting in my country, but I have noticed the amount of mass murders in the US caused by gun violence and would definitely help improve it.

Task:

The input for the algorithm can be a number of features such as: Incident Date, Incident Address, Number of Guns, Gun Type, Shooter Age, Shooter Gender, Shooter Relationship, Shooter Address, Number of People Killed, Number of People Harmed, Incident Description

The output for the algorithm could be used to answer several queries: - The most used Gun by the Shooters so that the access of the gun could be restricted further - It could be used to determine if there are some patterns with the same group of shooters and even used to determine the next location for the Incident - The reason behind most of the Incidents happening and then try to avoid them - Review the number of People Killed and Harmed and then try to figure out how to save more of them - Any connections between the Shooter and the Incident or the Shooter and the People or even the Shooter and the Address

However, one particular goal that I could find myself solving in this incident is the number of such incidents in a particular state in the next year.

The Goal in general is to reduce the amount of Gun Violence happening in the country

Data Preparation:

Collection of Dataset: The data can either be collected by getting a detailed survey from any police station and even contacting the https://www.gunviolencearchive.org for some estimated details regarding the topic. It can also be collected from an existing dataset from websites like kaggle, UCI etc. For example: https://www.kaggle.com/jameslko/gun-violence-data

Splitting the Dataset: To get the training set, testing set and the validation set; the entire dataset could be split into 3 parts. Random 70% of the entire dataset could be used to train the model as it is one of the most primary steps for the model to learn. The remaining 30% of the dataset could be used for the testing and validation phase. 15% of the dataset could be used for the testing set and the remaining 15% could be used as the validation phase.

Ground-Truth Labels: The ground truth or the labels for each training sample are provided with the dataset. We can then determine the error rates from root mean squared error method. We then improve the accuracy of the model by methods such as cross validation. We need to get the accuracy of the model to be as close to the ground truth as possible.