Ahmedabad University

CSE 523 Machine Learning
Weekly Report Progress

Topic - Password strength checker

Group name - Predictors_4.0

NAME	ENROLLMENT NUMBER
Zalak Shah	AU2040217
Freyal Shah	AU2040228
Freya Modi	AU2040222
Avinash Raval	AU2040031

After our Mid sem presentation, we tried to implement the changes/ suggestions made by the professor. Thus we tuned our efforts towards more rigorous work after the presentation. We re-submitted our Mid-Term Report with the necessary changes required in it.

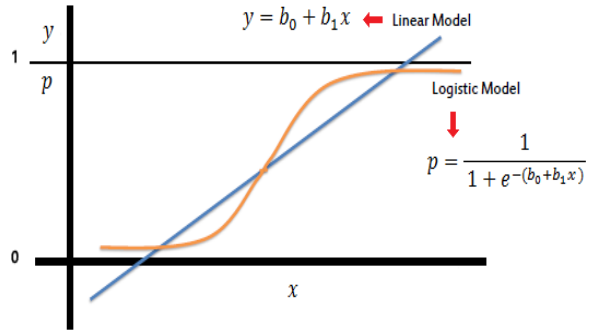
Edited Report :

https://docs.google.com/document/d/1klgMm7vvjmvEbB8_VO63ZzqSazbptM0O/edit?usp=share_link&ouid=115239433924492277168&rtpof=true&sd=true

Tasks performed this week:

1. Understanding Logistic Regression(Theoretical) :

It is a statistical method used for binary classification tasks in machine learning. It is a type of supervised learning algorithm that predicts the probability of an input belonging to a particular class (usually either 0 or 1).



In logistic regression, the input variables (also called features or predictors) are combined linearly to obtain a weighted sum, which is then passed through a logistic function (also known as sigmoid function) to produce the output. The sigmoid function maps any input value to a value between 0 and 1, which can be interpreted as the probability of the input belonging to the positive class.

The logistic regression model is trained by minimizing a cost function, which measures the difference between the predicted probability and the actual class labels.

2. Search for code :

We are using Scikit's in-built logistic regression model which is one of the best tools to kickstart .We try to define each and every necessity (Sigmoid function, Cost function, Gradient Descent , F1 Score) for better presentation of our work.

The formula gives the cost function for the logistic regression.

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$$

Where hx = is the sigmoid function we used earlier.

python code:

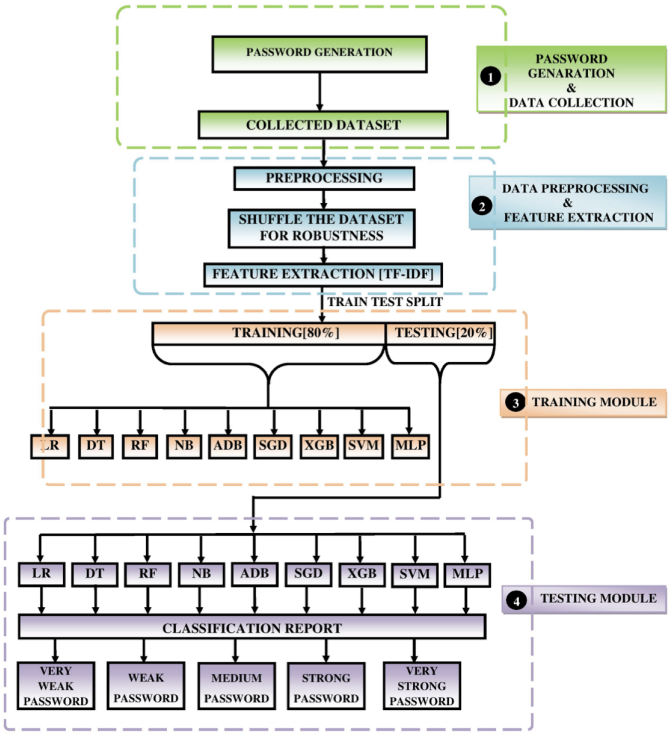
```
def cost(theta):
    z = dot(X,theta)
    cost0 = y.T.dot(log(self.sigmoid(z)))
    cost1 = (1-y).T.dot(log(1-self.sigmoid(z)))
    cost = -((cost1 + cost0))/len(y)
    return cost
```

The picture depicts the cost function and python implementation for the same..

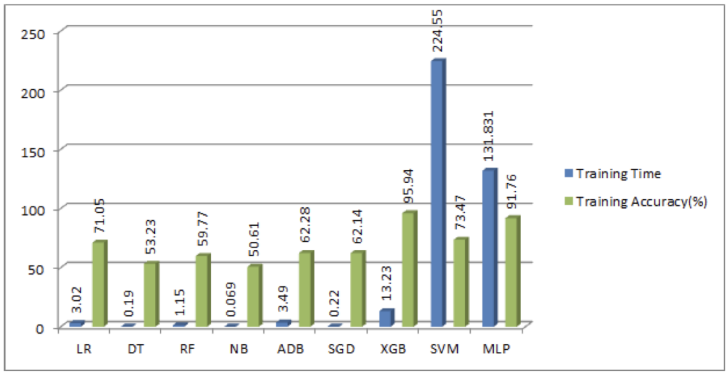
Reference:

https://www.analyticsvidhya.com/blog/2022/02/implementing-logistic-regression-from-scratch-using-python/

3. Futuristic Goals :



As suggested by the professor, we will try to broaden our prediction in terms of (Weak, Medium, Strong). The following chart depicts various ML models in use for a particular dataset and a single problem(password strength prediction).



The following bar chart represents the Training time and Accuracy for different models . This is definitely a better way to derive a comparative analysis for our problem. Thus we intend to get these results .

Reference :
https://assets.researchsquare.com/files/rs-1820885/v1_covered.pdf?c=1657729897