School of Engineering and Applied Science
CSE523- Machine Learning

Project Report - 3

Vehicle Insurance Predictor

Submitted to: **Prof. Mehul Raval**
Date of submission**: 2nd Mar 2021**

Members:

Name	Roll Numbers	Branch
Samarth Shah	AU1841145	B. Tech, ICT
Priyank Sangani	AU1841136	B. Tech, ICT
Yash Patel	AU1841141	B. Tech, ICT
Shaili Gandhi	AU1841012	B. Tech, ICT

Task performed this week:

- Understanding of hypothesis function of logistic regression.
- Understanding how sigmoid function works
- Loss function of logistic regression and how we can use gradient descent to minimize loss and get our model trained.

Outcomes of the task performed:

- So we had not yet implemented this model from scratch and we had understood the mathematical equations.
- Starting from the hypothesis function of linear regression:

$$h_{\Theta}(x) = \Theta^{T} X$$

- But linear regression will lead to any predicted value between $(-\infty, \infty)$ and to predict a categorical or output as class we need to map the value to $[0, 1]$.
- Here, we have a classification problem where we want to predict whether a customer would buy the insurance(0) or not(1).
- For using linear regression, we would have to set a threshold based on which the classification would be done. Suppose the threshold is 0.5 for predicting Let's say that a customer actually does not buy the insurance, and the predicted continuous value is 0.4. Now, this would give us the false result that the customer would buy the insurance.

Logistic Regression:

Output : 0 or 1

Hypothesis:

$$Z = WX + B$$

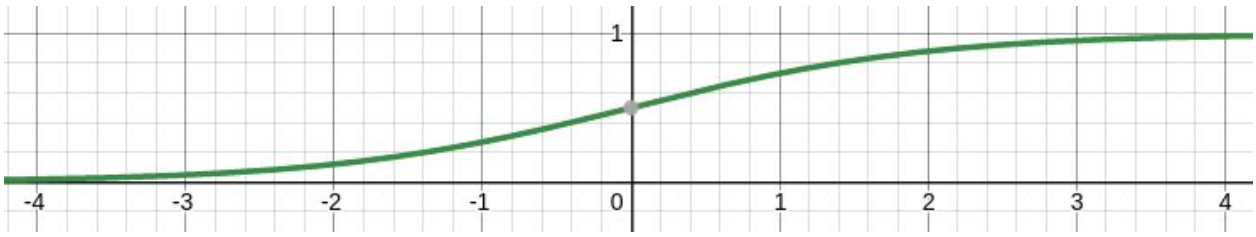
$$h\Theta(x) = \text{sigmoid}(Z)$$

For logistic regression, we can use the following logistic function,

$$h_{\Theta}(x) = 1 \div (1 + e^{(-Z)})$$

This function is also known as sigmoid function.

Graph of sigmoid function:



We get the estimated probability from the output of the hypothesis and this helps us to infer how confident can the predicted value be the actual value for an input X.
So in logistic regression, we try to fit the data in a linear regression model and then it is acted by a logistic function to predict the target categorical dependent variable.

Cost function:

$$\text{Cost}(h_\Theta(x), y) = -y \log(h_\Theta(x)) - (1-y) \log(1 - h_\Theta(x))$$

If y = 1, (1-y) term will become zero, therefore $-\log(h_\Theta(x))$ alone will be present

If y = 0, (y) term will become zero, therefore $-\log(1 - h_\Theta(x))$ alone will be present

Tasks to be performed in the upcoming week:

1) As we are done with mathematical intuition we would like to work on a python script to develop a logistic regression model for vehicle insurance classification.
2) Also we would like to search and understand the mathematical intuition of Support Vector Machine classification models.