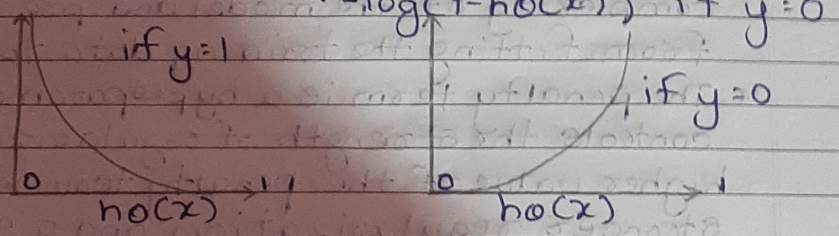
Assignment No: 3

Q.1 What is the cost function used in logistic regression and why it necessary?

1. The cost function represents optimization objectives ie we create a cost function and minimize it so that we can develop accurate model with minimum error.
2. If we try to use the cost function of linear regression in logistic regression then it would be of no use as it would end up being a non convex function with many local minimums, in which it would be very difficult to minimize the cost value and find the global minimum.
3. The cost function for logistic regression is

$$\text{cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$



above to equation can be composed in the single function ie

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

Q.2 Explain the concept of regularization in logistic regression.

1. Regularization is a technique used to avoid overfitting in machine learning models. It does this by adding a penalty term to the objective function called loss function or error function that the model is trying to minimize.
2. The objective function measures the error or difference between the predicted output of the model and true output. In logistic regression the objective function is typically the cross entropy loss which measures the difference between the predicted probability of positive class and the true label.
3. By adding a penalty term to objective function, regularization helps to reduce the complexity of the model and prevent it from fitting the training data too closely.
4. The penalty term is a hyperparameter that controls the strength of regularization. A higher value of the penalty term leads to stronger regularization and a simpler model, while lower value allows the model to be more complex.
5. It helps in reducing the complexity of the model by forcing the coefficients to be small.
6. It helps for improving generalization by reducing the variance of the model.

Q 3 What are the performance metrics commonly used to evaluate the performance of a logistic regression model in classification task ? Consider a spam detection example and calculate performace metrices for given values

TP = 85 TN = 10 FP = 5 FN = 0

Following are the commonly used performance metrices

1. Accuracy.

The accuracy metric is one of the simplest classification metrices to implement and it can be determined as the number of correct predictions to the total number of the predictions It can be formulated as.

$$\text{Accuracy} = \frac{\text{Number of correct Predictions}}{\text{Total number of predictions.}}$$

2. confusion matrix.

A confusion matrix is a tabular representation of prediction outcomes of any binary classifier which is used to describe the performance of the classification model on a set of test data when true values are known.

confusion matrix is simple to implement, but the terminologies used in the matrix might be confusing for beginners.

3. Precision

The precision metric is used to overcome the limitation of accuracy. The precision determines the proportion of positive prediction that was actually correct. It can be calculated as

the true positive or prediction that are actually true to the total positive predictions

$$\text{Precision} = \frac{TP}{TP+FP}$$

4. Recall

It is also similar to the precision metric however, it aims to calculate the proportion of actual positive that was identified incorrectly It can be calculated as True positive or prediction that are actuary true to the total number of positives, either correctly predicted as positive or incorrectly predicted as negative

$$\text{Recall} = \frac{TP}{TP+FN}$$

given

$TP = 85 \quad TN = 10 \quad FP = 5 \quad FN = 0$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$= \frac{85+10}{85+10+5}$$

$$= 0.95$$

$$\text{Precision} = \frac{TP}{TP+FP}$$

$$= \frac{85}{85+5} = 0.945$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{85}{85+0} = 1$$

Q 4 Explain the concept of multi class classification with logistic regression. How can it be achieved using technique one vs all ?

Q.5 What is support vector machine and what is it's primary objective in classification.

1. Support vector machine or SVM is one of the most popular supervised learning algorithm which is used for classification as well as regression problem.
2. The goal of SVM algorithm is to create the best line or decision boundry that can segregate n diamensional space into classes so that we can easily put the new data point in the correct category in the future This best decision boundry is called hyperplane.
3. SVM chooses the extreme points / vector that help in creating the hyperplane. These extreme cases are called support vectors and hence algorithm is termed as support vector machine.
4. SVM algorithm can be used for face detection image classification, text categorization etc.
5. Two types of SVM are their Linear SVM and non linear SVM.
6. Linear SVM is used for linearly separable data which means if a dataset can be classified into two classes by using a single straight line then such data is termed as linearly seperable and classifier is used called as linear SVM classifier.
7. Non linear SVM is used for non linearly seperated data which means if dataset cannot be classified by using a straight line, then such data is termed as non linear data and

classifier is used as called as non linear svm classifier.

8. The primary objective of svm in classification is to find the hyperplane that maximizes the margin while minimizing classification errors. This makes svm particularly effective for tasks where a clear seperation between classes is desired and they are known for their ability to handle high diamensional data and complex decision boundries.

Q.6 Explain the structures of a basic feedforward neural network including input, hidden and output layers.

1. Feed forward neural network also known as deep feedforward network or multi layer perceptrons.
2. Feed forward neural network is a key component of this fantastic technology since it, aids software developers with pattern recognition and classification, non linear regression and function approximation.
3. A feed forward neural is a type of artificial neural network in which nodes connections do not form a loop.
4. The data enters the input nodes, travels through the hidden layers and eventuaily exists the output nodes. The network is devoid of links that would allow the information exiting

the output node to be sent back into the network.

5. The purpose of feedward neural network is to approximate functions.

6. Feedforward neural network has following layers

a. Layer of input
It contains the neurons that receive input. The data is subsequently passed on to the next tier. The input layer's total number of neurons is equal to the number of variable in the dataset.

b. Hidden layers
This is the intermediate layer which is concealed between the input and output layers. This layer has a large number of neurons that perform alterations on the inputs. They can communicate with output layer.

c. Output layer
It is the last layer and is depending on the model's construction. Additionally the output layer is expected feature as you are aware of the desired outcome.

Q7 What is backpropagation and how does it work in training of neural network?

1. Backpropagation is the essence of neural network training. It is method of fine tuning the weights of neural network based on the error rate obtained in the previous epoch

Proper tuning of the weights allows you to reduce error rates and make the model reliable by increasing its generalization.

2. Backpropagation in neural network is a short form for backword propagation of errors. It is a standared method of training artificial neural networks. This method helps calculate the gradient of loss function with respect to all the weights in the network.

Working.

1. It begins with a forward pass where input data is processed through the network to make predictions followed by error calculation using a loss function to measure the prediction accuracy.
2. In the backword pass, the error gradient is computed and propagated from the output layer back through the network layers using the chain rule determining how much each neuron's output contribution to the error. This gradient information is then used to iteratively update the networks weights and biases nudging them in the direction that reduces the error.
3. This process repeats for multiple epochs until model converges to a desired level of accuracy.