CSCI 635: Introduction to Machine Learning

Homework 3: Multiclass Classification

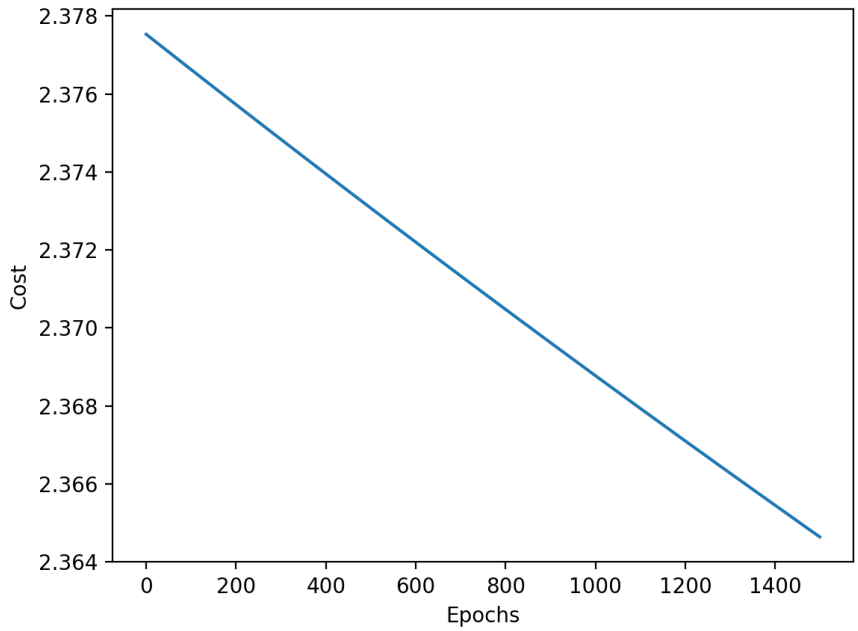
Question 1a

To tackle the XOR problem, first, the data was trained with the maximum entropy model. To calculate the model parameters, partial derivatives need to be computed which was done using gradient descent method.

The hyperparameters chose by me was a learning rate of 10^{-4}, weights and bias which is randomly initialized and an epoch of 200. The learning rate was chosen to be that particular value to tell how far to move the weights.

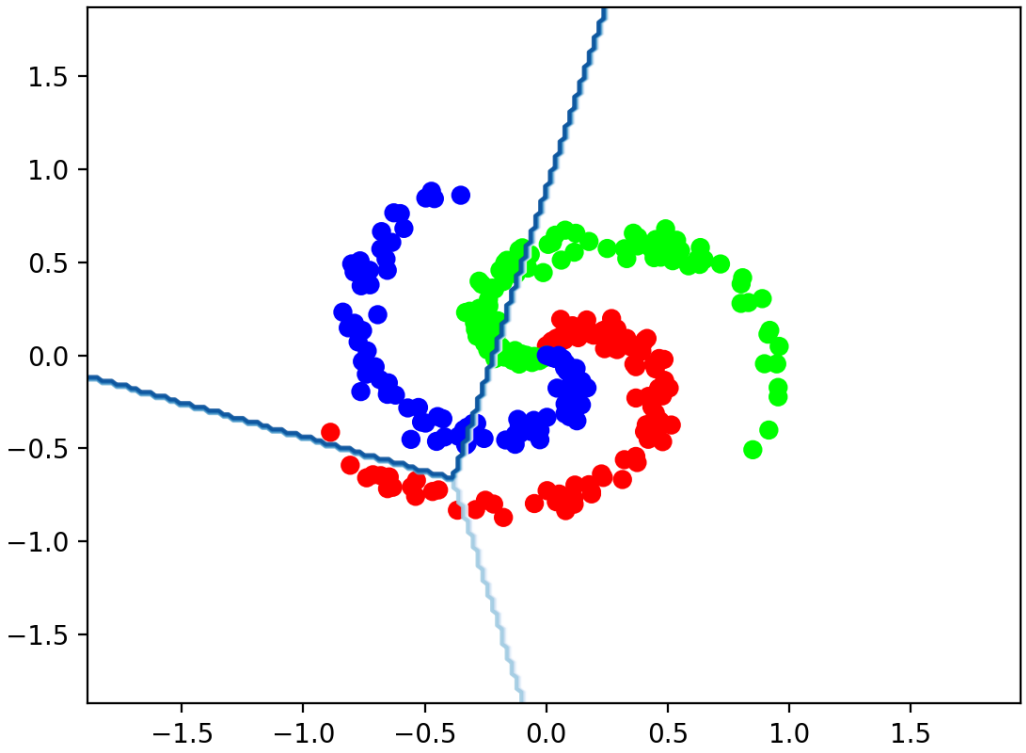
The model gave an accuracy of 50% which means the data fits properly with the model half of the time. It's loss per epoch reduces at a linear rate.

This is the cost function.



Question 1b

Decision boundary Plot:

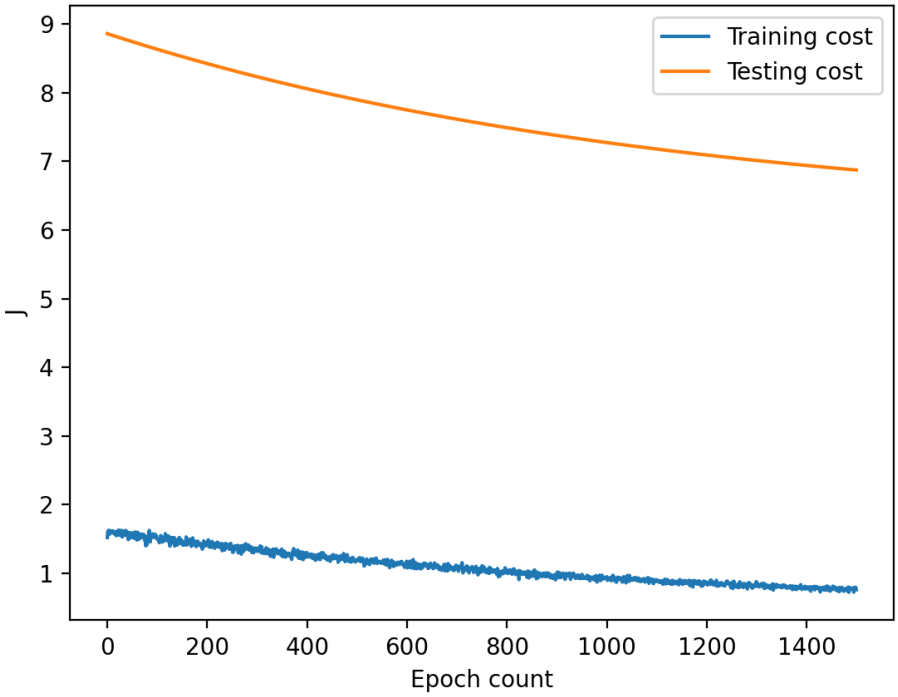


From the plot it is understood that the data does not exactly fit into the model. It fits almost fully correct in class but not exactly a perfect fit.

Accuracy : 46.03

Question 1c:

Training loss vs Testing loss:



The train accuracy = 27.2 % and test accuracy = 32.72 %

When the model is trained for more epochs, the error rate keeps decreasing and the accuracy increases for a while and then remains constant. If the epoch keeps increasing, it can eventually lead to overfitting of the data. Hence for this we must either use regularization or early stopping.

Question 2:

a)

The parameters computed for the model is the posteriori probability for the columns with categorical values. For the columns with real values, the mean and standard deviation is taken to be the parameter.

b)

The classification error rate achieved is 20.03%

c)

To perform classification based on chosen subsets, I would use a brute force approach and try each combination of features to see which combination would give the lowest error rate.