CS5590 – Deep Learning – Lab Assignment 1 Report

Logistic Regression for IRIS dataset using Tensorflow

Class Id 22: Avni Mehta

# Introduction

In the In-Class exercise, we learnt how to use Tensorflow to perform Linear Regression. However, when the response is qualitative, we cannot use linear regression. For this assignment, a logistic regression model is build using Tensorflow for Iris dataset. This dataset has four features and three labels. Given these four features, we have to predict flower labels.

# Objectives

For the given dataset, train a logistic regression model. When the four flower features are provided, our model should predict Iris flower species for this new data.

# Approaches / Methods

The concept for Logistic Regression is pretty similar to Linear Regression model. We still use the formula y = w. x + b, but with some tweaks.

Following are some of the changes made for Logistic model:

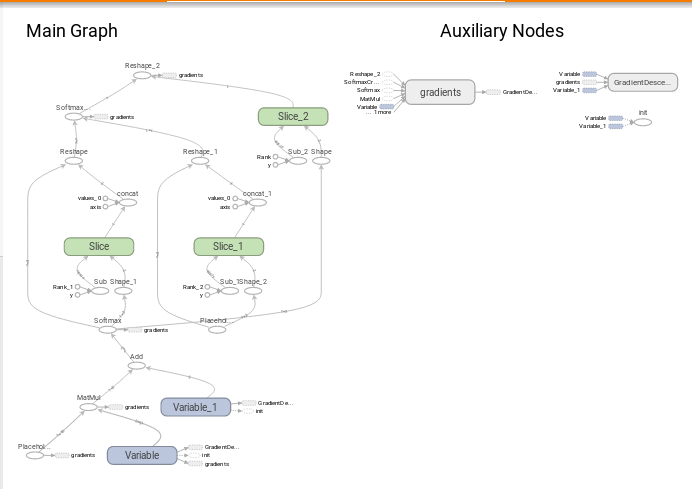
1. Use of softmax and cross entropy for loss function
2. Conversion to one-hot vectors (to encode categorical integer features)
3. Changing the label to numerical value

The steps involved in building logistic regression model for Iris dataset are:

1. Load the data from csv file.
2. Perform data-preprocessing:
3. Drop unwanted Id column
4. Convert qualitative response (species name) into numerical value
5. Convert the features and the labels to one-hot encoding.
6. Split the data into 80% training data and 20% testing data.
7. Define parameters for the model viz. learning rate and number of epochs.
8. Create placeholders for features(X) and labels(y). Create weights(w) and biases(b). Predict y\_ using X, w and b.
9. Define loss function using softmax cross entropy with logits.
10. Define training optimizer. Use gradient descent to minimize loss.
11. In the Tensorflow session, initialize all variables, write to tf.summary for generating graph.
12. For training the model, training data is used. Total loss for is calculated for every epoch by running the optimizer. The weight and bias is adjusted for every run to minimize the loss.
13. Test data is used for testing the newly built model. The predicted y-value is compared with actual y-value and the accuracy is displayed in the end.

# Workflow

Here is how the graph for logistic regression looks on Tensorboard:



# Datasets

The dataset used is Iris Flower Dataset. This data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: sepal length, sepal width, petal length and petal width, in cms. Based on the combination of these four features, the species is to be classified.

# Parameters

The features of the dataset are:

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm

The labels of the dataset are:

1. Iris Setosa
2. Iris Versicolour
3. Iris Virginica

The parameters used for building the model are:

* learning\_rate = 0.001
* num\_epochs = 150

# Evaluation & Discussion

When the learning rate is set as 0.001 for 150 epochs, the accuracy comes out to be 93.33%. But when the training data was tested for 15 epochs with the same learning rate, the accuracy was a bit lower, about 90%. The average loss decreased to approximately 0.4 between 150 epochs.

The accuracy for this model is pretty good as the dataset is clean and is meant for such classification problems.

# Conclusion

Linear regression is used to predict outcome whereas Logistic regression is used to classify the input based on the features. In this assignment we have adapted the formula for Linear regression after making some changes.

We achieved over 90% accuracy for this dataset with just 15 epochs and 0.001 learning rate. When the number of epochs were increased to 150, the accuracy improved to 93.33%. This dataset is very good for such classification problems.