Training Duration and Data Size Impact

Sailesh Dwivedy

METHODS:

In this Deep Learning experiment I used the sklearn dataset - Iris data [1], which is a multi class classification problem. Here we predict the target column where we have 3 classes - Iris setosa, Iris versicolor and iris virginica with 50 instances of each class. There are 150 records in total with 5 columns including the output/target column. There are is no missing values in any of the columns.

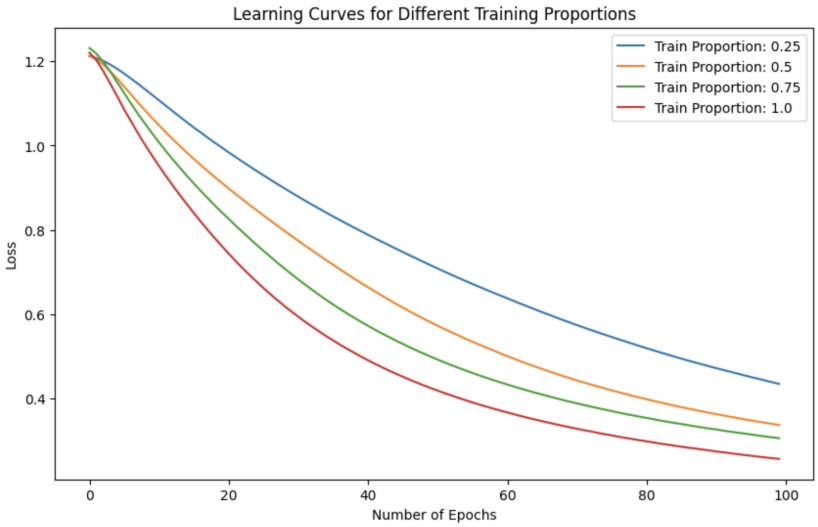
To train our model we created a 70/30 train test split of the dataset. Train comprises of 105 records and test comprises of 45 records. The input features were not on the same scale so we standardised the train(X_train) and test data (X_test) using a standard scaler because standardization helps to bring input features on same scale helping models learn better. I have used to random state as 42 for the experiment to ensure reproducible results.

I used MLPclassifier[2] as the model for the experiments. In this experiment I observed how training neural networks on different propertions of data affects the learning process keeping the number of hidden layers, number of neurons per layer, activation function, batch size, optimization approach and number of training epochs unchanged.

The 4 models have been trained on 25%, 50%, 75% and 100% of the data repectively. For all models the number of hidden layers is 2, neuron per layer is 20 and 30 respectively, activation function is relu, batch size is 21, optimization approach is stochastic gradient descent and no of training epochs 100.

Below we can see the learning process of each of the 4 neural networks trained on 25%, 50%, 75% and 100% of the data respectively.

RESULTS AND ANALYSIS:



From the above plot we can see the below trends:

1. Influence of amount of training data:

a. More training data leads to lower loss. We can see that when 100% of the data is used for training it has the lowest loss for the given no of epochs of 100.

Reason: This could be because of the fact that as model gets more data to train upon, it is able to identify patterns better and learn more about the data.

b. There is a diminishing trend in loss reduction as more data is used for training. The difference in loss reduction between 50% - 75%, 75% - 100% is less pronounced in comparison to the loss between 25% -50%.

Reason: This might be due to the fact that the once we have sufficient data to explain the uderlying patterns in the data, adding more data leads to only small improvements.

2. Influence of training duration:

a. For all models the loss decreases as the number of epochs increase.

Reason: This indicates that the models are learning and improving over time.

b. The rate of improvement or slope of the curve is steeper at the beginning as flattens as the no of epochs increase.

Reason: This is because initially the model makes significant improvements in learning the patterns. But as it converges the improvements become small.