Joseph Annand

DSE6211

Final Project Preliminary Results

Initially, a feed-forward dense neural network is trained using three quarters of all observations from the ABC hotels data set. The model consists of two hidden layers and the output. The first hidden layer consists of 75 units and uses a rectified linear activation function (relu).The number of units was chosen to be approximately two-thirds of the number of features in the training set. The relu function is the default activation function for hidden layers in a multilayer perceptron, so the relu function is used for both hidden layers of the neural network. For the second hidden layer, the number of units was chosen to be approximately half the amount of the units in the previous layer. For the output layer, the number of units is equal to 1 because the aim of the neural network is to provide a single output for predicting the status of a given booking. For a binary classification problem, the sigmoidal activation function is used so that all outputs are between 0 and 1.

The optimizer for the model is RMSprop, which is an adaptive learning rate algorithm that helps reduce the amount of computational effort required to train the neural network compared to other optimizers. The loss function used for the neural network is binary cross-entropy; this loss function calculates the dissimilarity between true labels and predicted labels when the output is between 0 and 1. The metric used in training the model is accuracy to measure how well the model is performing in terms of predicting booking status.

In the first attempt to train the neural network, a batch size of 512 and 100 epochs were used. The training and validation set results are shown in Figure 1.

A graph of different colored lines

Description automatically generated

Figure 1: Initial NN with batch size 512 and 100 epochs

There are over 27,000 samples in the training data set, allowing for about 54 batches per epoch in the first attempt at training the model. The model overfits the data as seen by the difference between the validation and training loss curves: the training loss continues to decline while the validation loss holds steady after about 40 epochs. Though this is a good sign that the model is well optimized, there is still room for improvement. The initial goal in training is to achieve a model that has some generalization power, or ability to make predictions on never-before-seen observations, and that can overfit the training the data. Figure 1 shows that the initial neural network created here does not sufficiently overfit, as the validation loss holds steady rather than reaching a minimum and then rapidly increasing.

In attempt to force the desired overfitting, the batch size is increased to 1000. The results are shown in Figure 2.

A graph of red and blue dots

Description automatically generated

Figure 2: Initial NN with batch size 1000

Increasing the batch size to 1000 did not help with optimizing the model. The model is still overfitting, but not to the point where the validation loss begins to trend upward. In order to achieve the best optimization and the most overfitting with the current optimizer, the representation power of the model must be expanded.

Two new models were created: one with three hidden layers and one with four hidden layers. Each additional hidden layer is the same as the first hidden layer of the initial neural network consisting of 75 units and using the relu activation function. Figures 3 and 4 show the learning curves for the models.

Increasing the size of the neural network does not appear to further optimize the model. The validation loss curves for the three-layer and four-layer neural network are similar to the curve seen in Figure 1. To further investigate the optimization of the initial neural network, the model using the initial network architecture of two hidden layers was fitted using a batch size equal to 512 and 200 epochs. The result curves are shown in Figure 5.

A graph of red and blue lines

Description automatically generated

Figure 3: Neural Network with three hidden layers

A graph of red and blue lines

Description automatically generated

Figure 4: Neural Network with four hidden layers

A graph showing different colored lines

Description automatically generated

Figure 5: Two-layer neural network using 200 epochs

When epochs is set to 200, the validation loss noticeably begins to trend upward and drastically diverge from the training loss curve, indicating the overfitting in the model desired to maximize optimization. Iterating over the entire training set 200 times is not practical, however, so additional changes to the network will have to be made to optimize the model more efficiently. The optimizer could be changed to an Adam optimization algorithm, and the learning rate could be adjusted to speed up training.

Additional data processing will need to be performed before evaluating the model using the test set. In the initial processing steps, after onehot encoding the test and training sets, the test set had less features than the training set. This is likely due to the training set having observations with rare values that are not seen in the test set. Identifying and either manipulating or removing the observations with these rare values will be a priority before the final report.