Lab2

1. Introduction

In this lab, we analyzed multiple datasets: the spam dataset and NBA dataset. These datasets are freely available from Kaggle.com.

1. Objectives

The objective was to analyze the datasets and fit appropriately accurate models. We also wanted to not only find models that accurately represented the data, but also did so in an efficient manner. We used loss and accuracy as the defining metrics as to how well the model fit the trained on the data.

1. Approaches/Methods

In this lab, we used several different methods of supervised and unsupervised learning. We used multiple linear regression, logistic regression, and two neural network models, including convolutional neural networks and long short-term memory networks.

1. Datasets

Here are listed the datasets that were used in our lab:

<https://www.kaggle.com/uciml/sms-spam-collection-dataset> - contains sms data

<https://www.kaggle.com/dansbecker/nba-shot-logs>

1. Conclusion

The datasets were all relatively clean, so getting good results from the required models was straightforward. We achieved an accuracy of 86.30% on the spam dataset using the convolutional neural network, and an accuracy of 97.00% using the long short-term memory network. Our results show how using long short-term memory works very well on natural language processing problems.