
# Report Assignment#2

# Forest Cover Type Classification

**Submitted to:**
Dr. Ayyaz Hussain

**Submitted By:**
Muhammad Tanzeel Saleem
04071913026

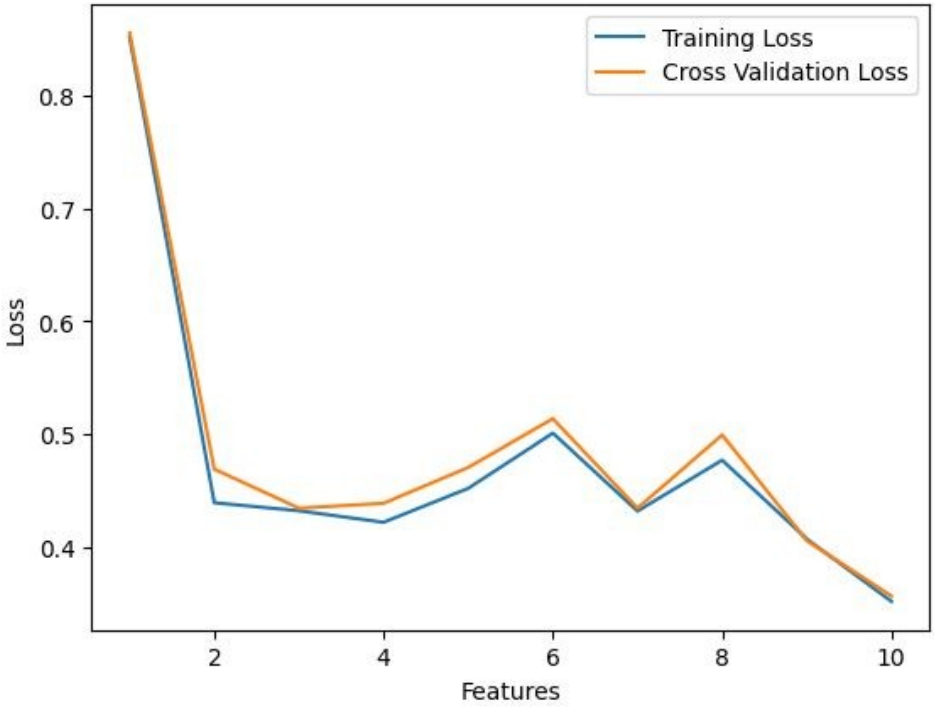
## Description of Data Set:

Predicting forest cover type from cartographic variables only (no remotely sensed data). The actual forest cover type for a given observation (30 x 30 meter cell) was determined from US Forest Service (USFS) Region 2 Resource Information System (RIS) data. Independent variables were derived from data originally obtained from US Geological Survey (USGS) and USFS data. Data is in raw form (not scaled) and contains binary (0 or 1) columns of data for qualitative independent variables (wilderness areas and soil types.

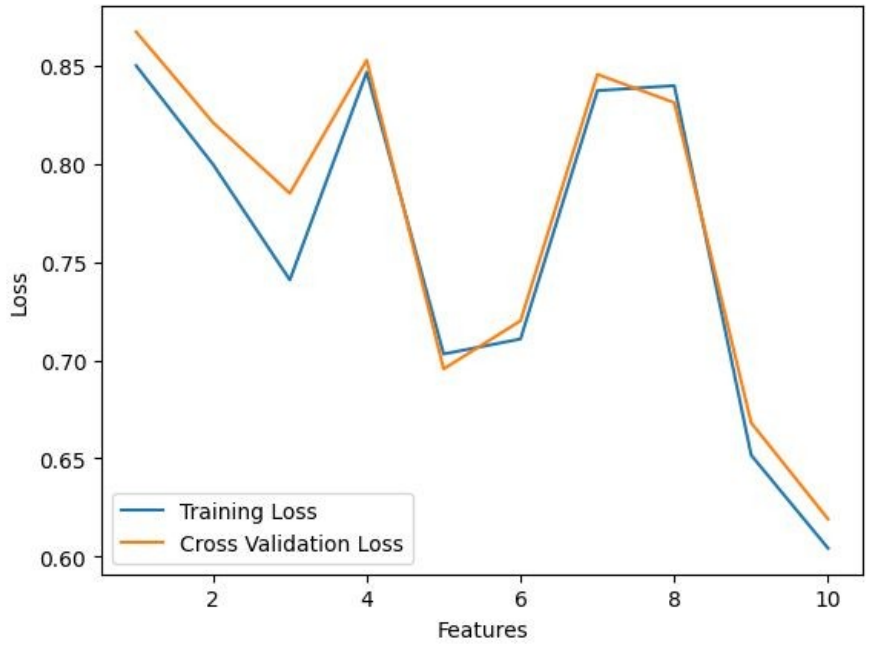
## Data Processing:

Data preprocessing encompasses fundamental procedures like ignoring textual attributes, removing missing values, applying under-sampling techniques to balance the dataset, and scaling the features.

## Analysis:

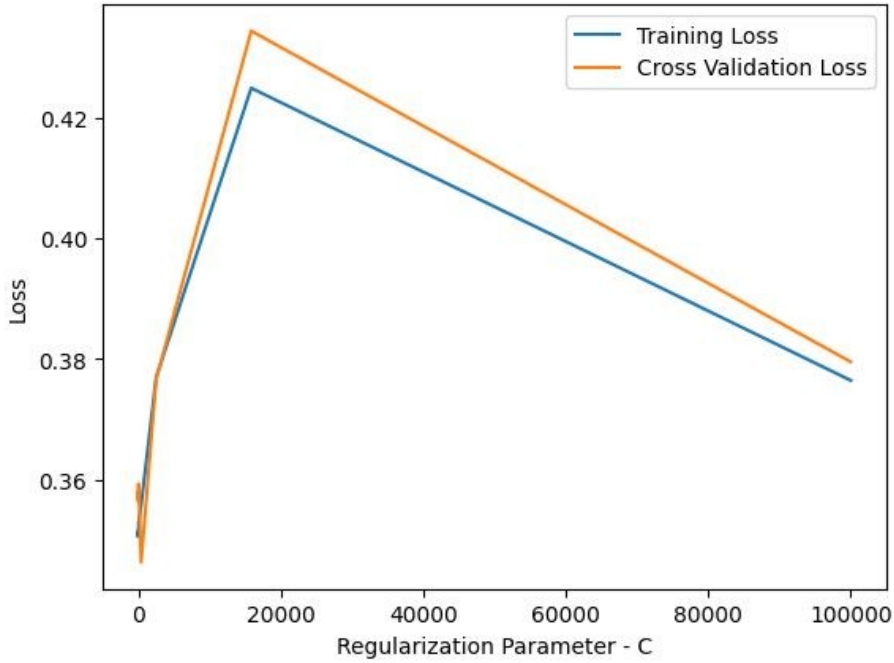



## -Logistic Regression

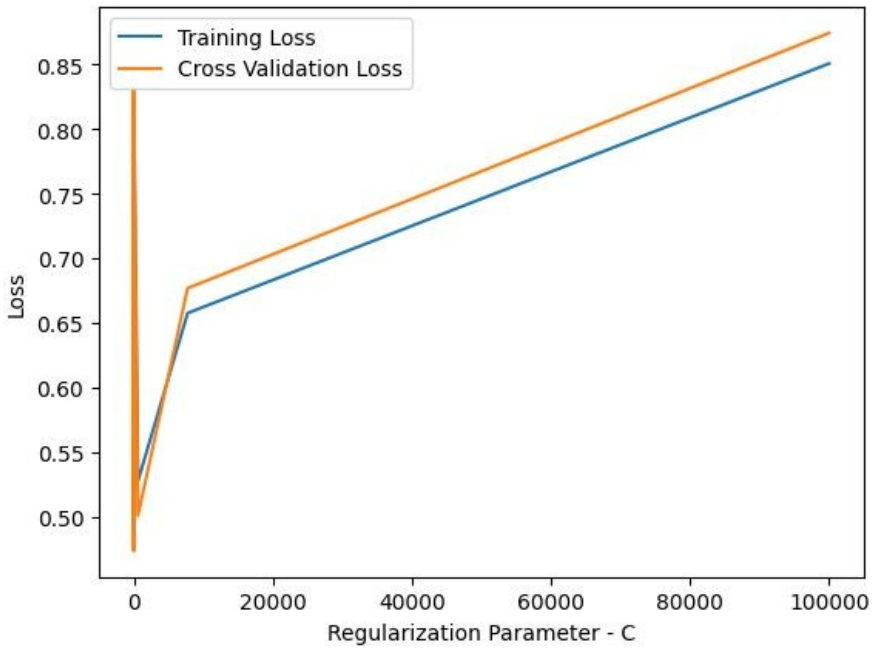



## -Neural Network

The graphs show that when we have fewer features in our dataset, the model tends to struggle with a problem called high bias or under-fitting. This means that the model is not capturing the complexity of the data and its predictions are too simplistic. However, when we include more features in our dataset, it helps alleviate the problem of high bias. By adding more features, the model has access to additional information and can learn more patterns and relationships within the data. This allows it to make better predictions and reduce the issue of under-fitting.

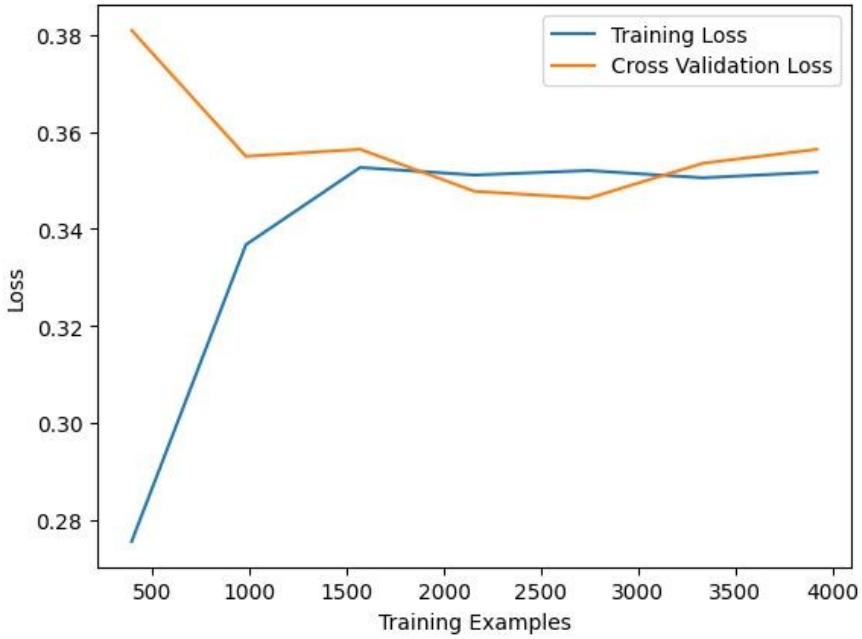



## -Logistic Regression

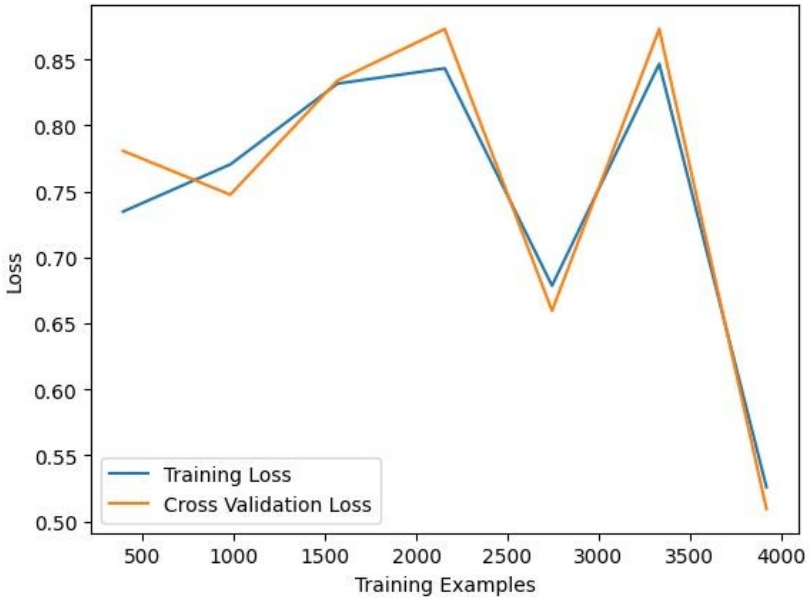



## -Neural Network

The graphs show that when we use a large value for the regularization parameter, the model tends to have a problem called high bias or under-fitting. This means that the model is not able to capture the complexity of the data and ends up making overly simplified predictions. However, if we decrease the value of the regularization parameter, it helps address the issue of high bias. By reducing the regularization, the model becomes more flexible and can learn from the data in a more nuanced way. This helps it make better predictions and avoid the problem of under-fitting.




## -Logistic Regression




## -Neural Network

The graphs indicate that when we have a smaller number of training examples, the model tends to struggle with a problem called high variance. This means that the model may not generalize well to unseen data and may overfit the training data. However, when we add more training examples, it helps reduce the problem of high variance. This allows the model to learn more patterns and improve its ability to make accurate predictions on new, unseen data. In simpler terms, having more training examples helps the model become more reliable and perform better on real-world scenarios.