Organic Compound Classification using Artificial Neural Networks

Utkarsh Sharma
skhiearth@gmail.com
linkedin.com/in/skhiearth

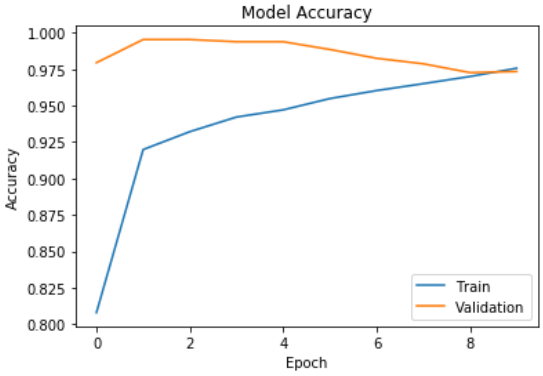
The given dataset contains details about organic chemical compounds including their chemical features, isomeric conformation, names and the classes in which they are classified. The compounds are classified as either 'Musk' or 'Non-Musk' compounds. The task was to classify these compounds accordingly. I used an Artificial Neural Network (Multi-Layer Perceptron) built using Keras to do the classification task.

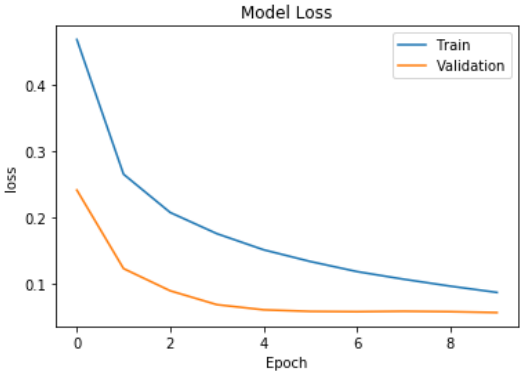
The structure of the neural network was as follows:

- Input Layer - 166 Features
- Hidden Layer - 20 Neurons (Activation Function - Tangent Hyperbolic)
- Output Later - 1 Output Feature - Class Label (Activation Function - Sigmoid)
- Optimizer - Adam
- Loss - Binary Cross-entropy
- Epochs - 10

The metrics of the model are as follows:

- Validation Loss: 0.055
- Validation Accuracy: 97.348%
- F1-Score: 0.989
- Precision: 0.991
- Recall: 0.988





The number of neurons and activation function for the hidden layer were determined using a GridSearchCV (the code for which was later removed from the final build).

I also performed Feature Scaling on the input dataset because the variance in the metrics was very high and could have caused the weights of the hidden layers to be biased towards one metric more than others.

The model training was performed in under 10 seconds on my local build.
Build Configuration: MacBook Pro (13-inch, 2017), 2.3 GHz Dual-Core Intel Core i5 with 8 GB 2133 MHz LPDDR3 memory.