Mount Rainer: Disappointment Trail, Success Rate Prediction using Linear Regression, Polynomial Regression and Multi-Layer Perceptron

Menaka Kollu
Arizona State University
mkollu@asu.edu

Nishanth Solomon
Arizona State University
nsolomo2@asu.edu

Sushilkumar Muralikumar
Arizona State University
smural32@asu.edu

***Abstract-The goal of this* Assignment is to work with the Mount Rainer Success rate dataset and predict the success rate of reaching the Disappointment Cleaver based on three features: temperature, humidity and wind speed. This is done using three methods: 1. Linear Regression 2. Polynomial Regression and 3. Multi-Layer Perceptron Model(MLP). The training RMSE and the Testing RMSE for the different models have been recorded and discussed.**

I. Problem Formulation

The goal of this group task is to predict the success rate of a person climbing the Disappointment Cleaver summit based on the weather conditions prevalent during that day.
This is performed using three models and they are as follows:

- Linear Regression
- Polynomial Regression
- MLP

The Root Mean Square Error(RMSE) for the testing and training phases of the models have been recorded and the results have been discussed.
This task has been performed using the following steps

1. Preparation of data and choosing features
2. Preparation of data and dividing into training and testing dataset
3. Building linear regression model
4. Calculating RMSE of linear regression model
5. Creating a Polynomial Regression Model.
6. Calculating RMSE of the polynomial regression model.
7. Training a multilayer perceptron(MLP)
8. Calculating RMSE of MLP model
9. Adjusting the parameters of MLP to achieve comparable performance as linear regression model in terms of RMSE

II. Methods

Preparation of Data:
The provided dataset contained the data pertaining to the Number of attempts and the Number of success trials at various trails at Mount Rainer, on different days of the year for a period of two years(2014 and 2015). The dataset also provided a record of the weather conditions recorded at the area on any given day during the two year period. The goal of this task is to predict the success rate of a person taking the Disappointment Cleaver trail on any given day. For this prediction the following weather conditions were considered:

- Temperature
- Humidity
- Wind-Speed

Preprocessing of the data: 2 csv files climbing_statistics and Rainier_Weather are considered. It has been observed that the *date* in climbing_statistics are repeating , so the *groupby* function of pandas is useds and *sum* the data corresponding to the given *date* . We have merged both the datas and have used forward and backward interpolation to fill the missing weather conditions for a given date in climbing statistics data. FinalData_all.csv has the final dataset used for the models.
The size of the dataset is 331*6(where the first column contains information about the date on which the data was recorded).

The data was split for training and testing of the models. The split was 80-20, where 80% of the data is used to train the models and the remaining 20% is used for testing. The 80% training data was further split into 70-10, with 10% of the data being used for validation. The data was split randomly using Python thus ensuring that no single day contributes more to the testing phase, thus, ensuring proper functioning of the models.

The data after being split into training and testing datasets, were normalized using the Scikit Learn library's Normalize function.

Linear Regression Model:

In this model we are predicting the happiness score using the three selected features (*temperature, humidity and wind speed*) .
This is a Multivariate Linear Regression problem.
Hypothesis for this model is given by :

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n$$

where, n=3
and x is the input
We are using LinearRegression model provided by sklearn[1]

Polynomial Regression Model:

In this model we are predicting the happiness score using the five selected features (*gdp, life_expectancy, freedom, generosity, corruption and generosity*) .
This is a Multivariate Linear Regression problem.
Hypothesis for this model is given by :

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \cdots + \beta_n x^n + \varepsilon.$$

where, n=5
and x is the input
We are using LinearRegression model provided by sklearn[1]

Multilayer Perceptron model:

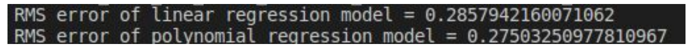
The goal is to formulate a Multi-Layer perceptron model which can perform on par compared to the Multivariate Linear Regression model used earlier. The Sklearn library's MLPRegressor[2] model has been used for this purpose.
A Multi-Layer Perceptron model with one hidden layer was chosen for this application. The model has three input neurons, 100 neurons in the hidden layer and one output neuron. This was arrived at based on different test cases run by varying the metrics used in the hidden layer and the execution of the program. The metrics that were varied are as follows:

- Number of neurons in the hidden layer
- Activation function
- solver
- batch_size

The file training.csv contains some of the test cases run by varying the metrics. The results of the analysis are as shown in results section

III. Results

Linear Regression Results:
The RMSE value for the linear regression model:


```
RMS error of linear regression model = 0.2857942160071062
RMS error of polynomial regression model = 0.27503250977810967
```
Figure 1: RMSE for linear regression model

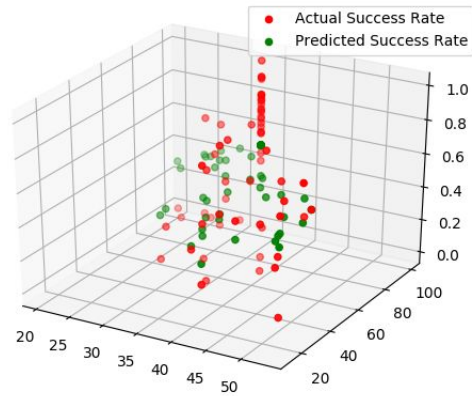


Figure 2: Plot between Success Rate, Temperature and Wind speed

Polynomial Regression Results:
The RMSE value for the linear regression model:


```
RMS error of polynomial regression model = 0.27503250977810967
```
Figure 3: RMSE for linear regression model

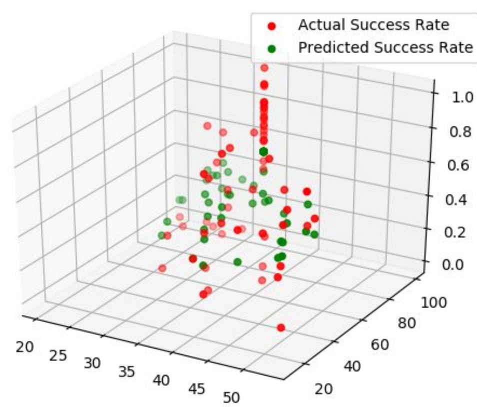


Figure 4: Plot between Success Rate, Temperature and Wind speed

Multi-layer Perceptron Results:
The RMSE value for the multi-layer perceptron :

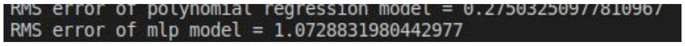

RMS error of polynomial regression model = 0.27503250977810967
RMS error of mlp model = 1.0728831980442977

Figure 3: RMSE for Multi-layer perceptron model

Parameters:

```
activation=relu
solver=lbfgs
max_iter=1000
batch_size=10
alpha=0.01
learning_rate=adaptive
learning_rate_init=0.001
shuffle=False
early_stopping=False
validation_fraction=0.1
n_iter_no_change=10
```

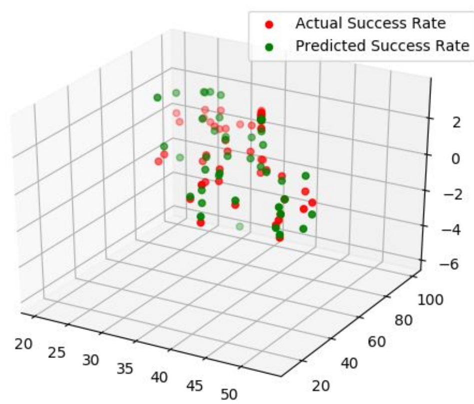


Figure 6: Plot between Success Rate, Temperature and Wind speed

IV. Conclusions

This task has focussed predicting the success rate of reaching the Disappointment Cleaver sumit given the weather conditions on a particular day by using three different models. The performance analysis of these three models was performed using the RMSE metric. The RMSE obtained for the Linear Regression model is 0.2857, for the Polynomial Regression is 0.2750 and the RMSE for the Multi-Layer Perceptron model was found to be 0.6168. It was also observed that the MLP model performed better when a different solver such as Low Memory Broyden-Fletcher-Goldfarb-Shanno(LBFGS) was used. Thus it was observed that the polynomial regression model was observed to work better.

References

[1]https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

[2]https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html#sklearn.neural_network.MLPRegressor