The better solution: A comparison of Linear Regression and Multi-Layer Perceptron based on RMSE

*

Omik Save
School of Energy Matter and Transport Engineering
Arizona State University
Tempe, United States
osave@asu.edu

Gautam Sharma
CIDSE
Arizona State University
Tempe, United States
gsharm17@asu.edu

Hrishikesh Patel
CIDSE
Arizona State University
Tempe, United States
hjpatel9@asu.edu

***Abstract*—This report discusses comparison of Multi-Layer perceptron (MLP), Linear Regression and Polynomial Regression for prediction of success rate of trails based on weather conditions. RMSE was chosen as the core parameter to validate the efficacy of all algorithms. This report also highlights strategies to fill missing datapoints of a key parameter, i.e. wind speed. The polynomial regression model was chosen to discuss interaction between individual features rather than including n_{th} order of the same parameters.**

***Index Terms*—MLP, linear regression, multi-variable regression, neuron, SGD, ReLU, ADAM solver.**

I. Data Preparation

The data was a combination of weather statistics on Mount Rainier provided by Northwest Avalanche Center and climbing statistics by various group provided by Mount Rainier Climbing source. The key association parameters were the temperature, humidity and wind speed of the weather with climbing statistics i.e. number of attempts of Disappointment Cleaver and total number of successful completions of the trail, thus the success rate is defined as the ratio of success and attempts. It was observed that the wind speed data had missing datapoints. Since, wind speed is one of the three parameters that determine the succeed percentage.

- Downloaded/imported both dataset .csv files. First dataset provides information of succeed ratio for the summit depending on the route and the date they have started. Second dataset provides the weather conditions on the Mount Rainier attributed as temperature, wind speed and humidity.
- We removed all datapoints for the different routes except Disappointment Cleaver summit.
- Total wind speed datapoints are 465. Among those, 378 datapoints were remained that had nonzero wind speed. We computed mean and Standard Deviation for remaining 378 points. We replaced zero values of the wind speed by random values between the range including half of the standard deviation from mean.
- We merged both datasets with Key Mapping. In both datasets, 'date' as a common key was used resulting in a dataset with weather condition parameters to the corresponding succeed rate.

II. Algorithm Implementation

While the goal of this research is minimising error in prediction, on a broader spectrum it also focuses on performance of three algorithms namely, Linear Regression, Polynomial Regression and Multi-Layer Perceptron.

A. Linear regression

A simple linear regression [2] finds constants for a given set of input parameters in order to match the instantaneous output. It uses the mathematical form of slope-intercept to produce predictions given by,

$$y = mx + c \tag{1}$$

where y is the instantaneous output, x is the instantaneous input and c is a constant to measure offset from origin. A modified equation based on simple regression allows fitting dissimilar number of outputs and inputs given by

$$y = w_n^T \times x_n \tag{2}$$

where w represents weights of features and n denotes number of features. The weights of 2 are generated using training data and later the model is implemented on test data to find accuracy of prediction. An error is defined as the difference between predicted and the loss value.

B. Polynomial Regression

While linear regression fits features x with outcomes y, it does so with feature order restricted to one. This necessarily fails to determine interaction between individual features. Polynomial Regression, a variation of linear regression develops model that takes into account individual features, their n_{th} order as well interaction between individual features i.e. $n_i \times \mathrm{n}_j$. This research explores the interaction between all three parameters mentioned above in the mathematical form given by,

$$h = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_i x_{1,2,3} \times \theta_j x_{1,2,3} \qquad (3)$$

The hypothesis states that the interaction between parameter dominates whether the weather condition at a random day suits successful trails. If observed that the co-efficient of interaction is very less, it can be determined that individual features are independent and contribute separately towards the success rate.

C. Multi-layer Perceptron

MLP is an multi-linked algorithm whose indivisible part is a single neuron. A neuron can be assumed as a junction that sends and receives data. For an artificial neural network with one hidden layer is given by,

$$y = \phi(b^{(2)} + W^{(2)}(\psi(b^{(1)} + W^{(1)}))) \qquad (4)$$

where b is network bias and W is weights between input to hidden layer and hidden to output layer respectively. ϕ and ψ are activation functions chosen based on application. The properties of MLP model is determined by ReLU [3] activation functions and ADAM solver [1]. A ReLU is a rectifier function that outputs 0 for $x \leq 0$ and x for $x > 0$. Similarly, the derivative of ReLU is 0 for $x \leq 0$ and 1 for $x > 0$. The ADAM solver features an adaptive learning rate as compared to constant learning rate for a stochastic gradient descent. Thus, ADAM updates learning rate per-parameter basis and achieves faster convergence even with noisy sparse data.

III. Result

A. Regression prediction

The linear regression prediction of this data resulted in a highest RMSE of 0.4510. The coefficients of predictions [0.425 0.00358, -0.000760, -0.0057] associated with temperature, relative humidity and average wind speed. Thus, it can be declared that increase in temperature is favoured for success of trail. It can be also observed that the increase in wind speed and humidity does not support the success of trail.
The reported RMSE for polynomial regression of degree 2 was 0.426. The associated factors namely temperature, humidity and wind speed as well as interaction of all three parameters was reported as [0.06144768 0.01723597 -0.04726455 0.02993295 0.00621388 0.00169976] respectively. This result gives unclear intuitions however i.e. increase in humidity favours the success. It also does not provide clear intuitions about interaction factors and hence it cannot determine how interaction of wind and humidity will favour success of trail. Thus, polynomial regression cannot be relied as a robust model prediction.

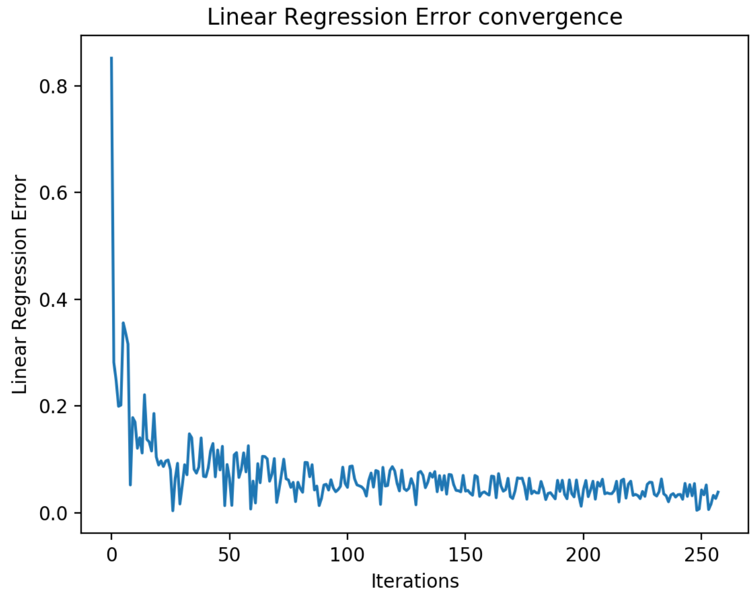

Fig. 1. Error convergence Linear Regression

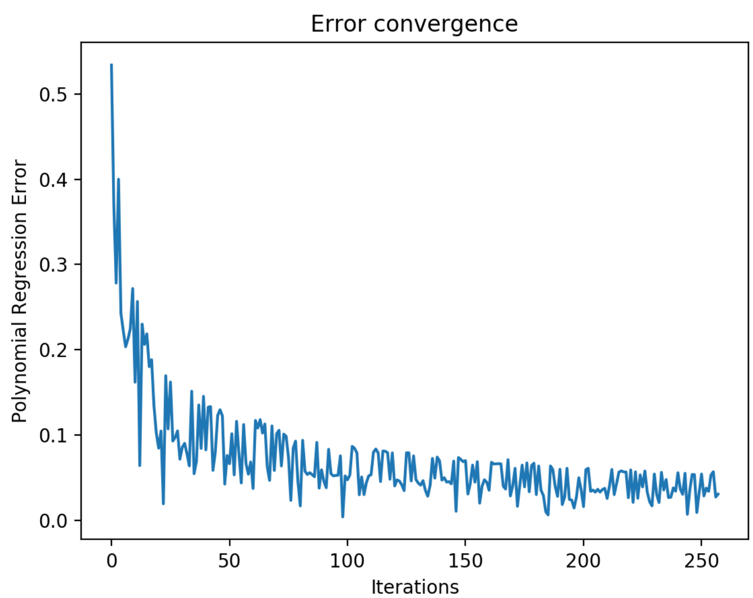

Fig. 2. Error convergence Polynomial Regression

B. MLP

The RMSE reported by MLP was 0.4506. This result was reported by implementing 3 input features, one output feature and one hidden layer with 8 neurons.

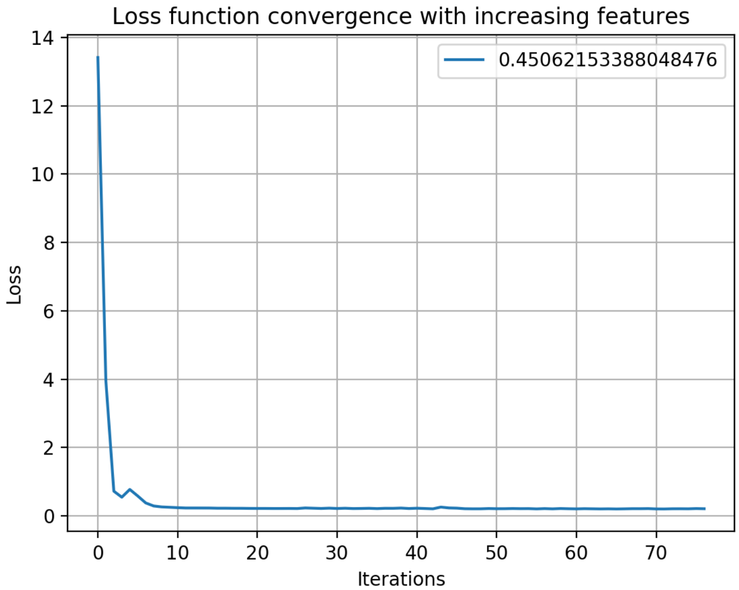

Fig. 3. Error convergence in Multi Layer Perceptron

IV. Conclusion

References

[1] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: 1412.6980 [cs.LG].

[2] Yale University. *Linear Regression*. 1997. URL: http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm.

[3] Bing Xu et al. *Empirical Evaluation of Rectified Activations in Convolutional Network*. 2015. arXiv: 1505.00853 [cs.LG].