ISSN:0975-9646

Weather Prediction Using Normal Equation Method and Linear regression Techniques

Sanyam Gupta[#1], Indumathy K[*2], Govind Singhal[#3],

[#] *Student, VIT University, Vellore, Tamil Nadu, India*

[*] *Assistant Professor, VIT University, Vellore, Tamil Nadu, India,*

***Abstract*— The burgeoning research in the fields of Artificial Intelligence and machine learning has given rise to numerous weather prediction models. But the problem of accurately predicting or forecasting the weather still persists. Numerical weather prediction is taking the existing numerical data on weather conditions and applying machine learning algorithms on it to forecast the weather. This paper is the application of machine learning algorithms, linear regression model from statistics, and two optimization techniques, Normal equation method and Gradient descent method to predict the weather on the basis of few parameters. Two optimization techniques have been used to compare the performance of the algorithms. The obtained results demonstrated that the normal equation method forecasts the weather with high precision, whereas the gradient descent method forecasts the weather with very little precision.**

***Keywords*—Weather prediction, machine learning, efficiency, linear-regression model, normal equation method, gradient descent method, dataset, R-studio, temperature, humidity, dew-point, error**

I. INTRODUCTION

Weather forecasting means predicting the weather conditions (conditions of atmosphere) of a particular given area or location. More importantly, accurate weather prediction is very important to pursue day-to-day activities. Living and non-living things are dependent on weather predictions.

Even after decades of weather forecasting, the weather industry in India is still in its initial stage, facing many obstacles. One of the major obstacles that weather forecasting faces is the arbitrary & ill-suited expectations from the nature.

Machine learning is the ability of computer to learn without being explicitly programmed. It allows machines to find hidden patterns and insights. In supervised learning, we build a model based on labeled training data. The model is then used for mapping new examples. So, based on the observed weather patterns from the past, a model can be built and used to predict the weather.

Numerical Weather Prediction (NWP) is a widely used prediction model. Mostly, it takes the present weather conditions and processes it to build a model for predicting the weather.

Several data mining techniques have been employed in diversified applications such as predicting rainfall, weather, storms and flood. Weather forecasting falls under predictive mining which focuses on the data analysis, formulates the database, and forecasts the features of anonymous data. [1]

This research work focuses on solving the weather prediction anomalies and in-efficiency based on linear regression algorithms through normal equation model. The major contribution of this research work is to formulate an efficient weather prediction model based on the linear regression algorithms. Thus, in order to overcome the limitations of weather prediction, this work uses normal equation model's hypothesis and compares it with the gradient descent model to provide a better idea of efficiency of the models.

Even though among the linear regression algorithms, gradient descent is considered as the widespread approach, this work utilizes normal equation algorithm to predict the future weather conditions with greater precision. With this research, weather can be forecasted with greater accuracy, which will be helpful in daily activities.

II. RELATED WORK (LITERATURE SURVEY)

In this section, an analysis is carried out with the existing weather prediction techniques available in the literature based on the diversified concepts of various researchers. Other weather prediction approaches based on linear regression method is presented for which various weather attributes have been used. Experimental result is calculated for different parameters, for each approach. In Sections III and IV, experimental results evaluated on dataset of weather conditions of the city Vellore, Tamil Nadu, which includes 10 years of meteorological data is discussed. The characteristics of the proposed system are discussed. The main characteristic of this research is that it is very efficient with the normal equation model. Especially, the addition of seasonal parameters improves the efficiency and performance of the proposed model.

Linear regression is the most basic and frequently used predictive model for analysis. Regression estimates are generally used to describe the data and elucidate relationship between one or more independent and dependent variables. Linear regression finds the best-fit through the points, graphically. The best-fit line through the points is known as the regression line.

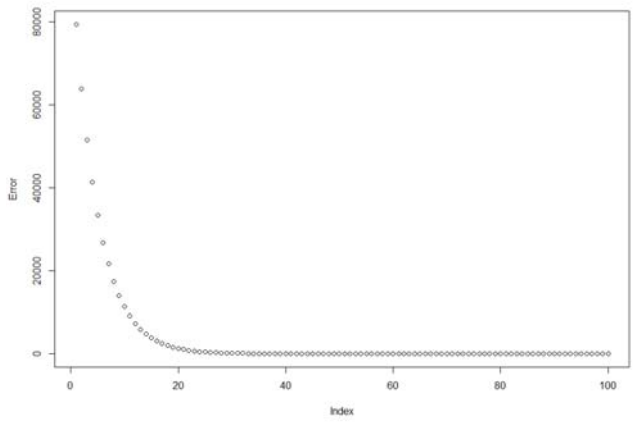

Fig. 1. Example of regression line

Fig. 1 is an example of the best-fit line. Here, the line can be straight or curved depending on the data. The best-fit line can also be a quadratic or polynomial which gives us better answer to our questions. Two of the optimization algorithms used in this research are Normal Equation Method & Gradient Descent Equation.

Independent Component Analysis (ICA) determines the independent component similarity within the spatial-temporal data. Here, neural networks together with nonlinear canonical correlation assessment are proposed for the purpose of determining the association among Humidity, Pressure, Temperature, wind speed and other weather parameters. [2] [3]

III. Methodology & Proposed Work

In a developing country and an economy like India where major population is dependent on agriculture, weather conditions play an important and vital role in economic growth of the overall nation. So, weather prediction should be more precise and accurate. Weather parameters are collected from the various stations of meteorological department, Tamil Nadu (Vellore). [2] The data used in this research is for the years 2014, 2015 and 2016. This research uses a software called 'R-studio' for programming. The programming language used is 'R-language'. Fig. 2 visualizes the system in the form of a block diagram.

Fig. 2. System block diagram

Two algorithms have been used to compare the efficiency when applied on the original dataset. Hypothesis for our proposed model can be shown by the Equation 1:

$$AvgTemp = a + b\ (AvgDew) + c\ (AvgHum) + d\ (AvgPress) + e\ (AvgVisi) + f\ (AvgWind) \quad (1)$$

Here, parameters a, b, c, d, e & f are known as the learning parameters. The optimization algorithm learns these parameters with the help of the training data fed to it.

Three types of weather parameters are predicted: Temperature, Humidity and Dew-point. Temperature is the measure of hotness or coldness, generally measured using thermometer. Units of temperature most frequently used are Celsius and Fahrenheit. Humidity is the quantity of water vapor present in the atmosphere. It is a relative quantity. Dew point is the temperature of the atmosphere (which varies according to pressure and humidity) below which water droplets begin to condense and dew is formed.

Table I shows sample data from the training set, a huge dataset is fed to the model. Parameters are trained and they correspond to the model. Finally, it can be said that the machine has been trained.

TABLE I
Sample Training Data

Date	Avg Temp	Avg Dew	Avg Hum	Avg Pres	Avg Visi	Avg Wind
1/1/2014	22	21	87	1016	3	0
2/1/2014	21.5	21	88	1016	3	0
3/1/2014	20.5	19	83	1017	3	3
4/1/2014	17.5	17	84	1017	3	10
5/1/2014	20.5	19	84	1015	3	6
6/1/2014	18	17	84	1014	3	0
7/1/2014	20.5	19	88	1014	3	0
8/1/2014	21	19	84	1016	3	0
9/1/2014	20	19	86	1016	3	0
1/1/2014	22	21	87	1016	3	0

The linear regression hypothesis is shown in Equation 2:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \ldots + \theta_n x_n \quad 2)$$

Least-square or the cost equation has been used in the algorithm of linear-regression method. The least-squares cost equation is shown in Equation 3:

$$J(\theta_{0\ldots n}) = \frac{1}{2n} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2 \quad (3)$$

The optimization algorithms used in the research are normal equation and gradient descent. Normal equation is an analytical solution to the linear regression problem with a least-square cost function.

The vector form of normal equation is shown in Equation 4:

$$\theta = (X^T X)^{-1} X^T y \quad (4)$$

Gradient Descent formula is shown in Equation 5:

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})\, x_j^{(i)} \quad (5)$$

R-language and Python is used to write the code. The results are obtained and weather entities are predicted using R-Studio. An application is published as well, that will help individuals to access it through a unique IP address. Anyone can use the application to look for weather conditions. The application is called Shiny App.

IV RESULT

The evaluation and simulation is carried out in R-studio. Prominently, three parameters: Temperature, Humidity and Dew-point are calculated through the hypothesis of the linear-regression model. The error is also calculated for each of the algorithms.

Parametric and total error is calculated. The results of the normal equation algorithm show very little deviation from the actual output. It can be deduced from the graphs and tables in the following sections that the error produced by gradient decent method is very significant compared to the error produced by normal equation method.

In the following figures, blue color shows the actual output, green color is the prediction by normal equation and red color shows the prediction by gradient descent method.

The results obtained for the entity **temperature** are shown in Table III and Fig. 4.

TABLE III
ERROR IN TEMPERATURE

Parameters	Normal Equation	Gradient Descent
a	154.3793	0.4349
b	0.900699	1.872743
c	-0.15097	-0.92408
d	-0.13633	0.050531
e	0.089188	0.59752
f	-0.12941	-0.72156
Error	0.599866	41.56697

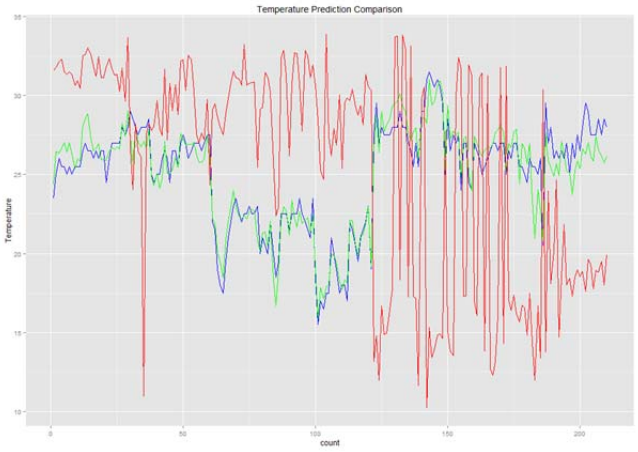

Fig. 4. Actual output (blue), prediction by normal equation (green) and prediction by gradient descent (red) for temperature

It can be deduced from the above graph that the difference between blue lines (actual output) & green lines (results by normal equation method) are very minimal, which signifies the precision and effectiveness of the normal equation method.

The results obtained for the entity **humidity** are shown in Table IV and Fig. 5.

TABLE IV
ERROR IN HUMIDITY

Parameters	Normal Equation	Gradient Descent
a	-1022.97	0.180542
b	-2.3754	-0.38342
c	4.065956	-1.14582
d	1.059731	0.094575
e	-0.43401	1.354212
f	-0.63411	0.717851
Error	16.05951	134.9208

Table IV shows the minimal error using normal equation and the large deviation from the actual output produced by gradient descent algorithm, showing the inefficiency of the gradient descent model.

Humidity plays a very significant role in predicting rainfall and droplets in atmosphere. It is also a deciding factor in many of our day to day activities. It also affects the temperature. As the relative humidity decreases, the temperature increases, and vice versa.

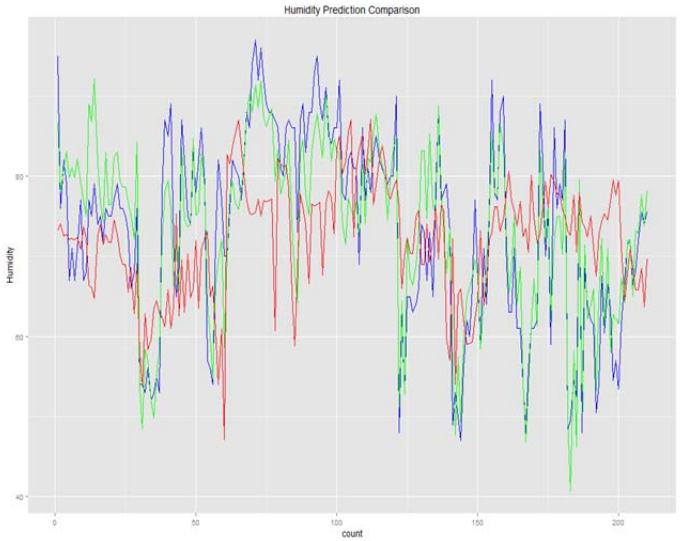

Fig. 5. Actual output, prediction by normal equation and prediction by gradient descent for humidity

It can be deduced from Fig. 5 that normal equation predicts more accurate results as compared to gradient descent method, even in the case of humidity.

The results obtained for the entity **dew-point** are shown in Table V and Fig. 6.

TABLE V
ERROR IN DEW-POINT

Parameters	Normal Equation	Gradient Descent
a	164.8517	-0.62643
b	0.469616	-1.18322
c	0.134735	0.48247
d	-0.16361	0.026811
e	0.077588	0.369279
f	0.077965	-2.42383
Error	0.740194	156.8855

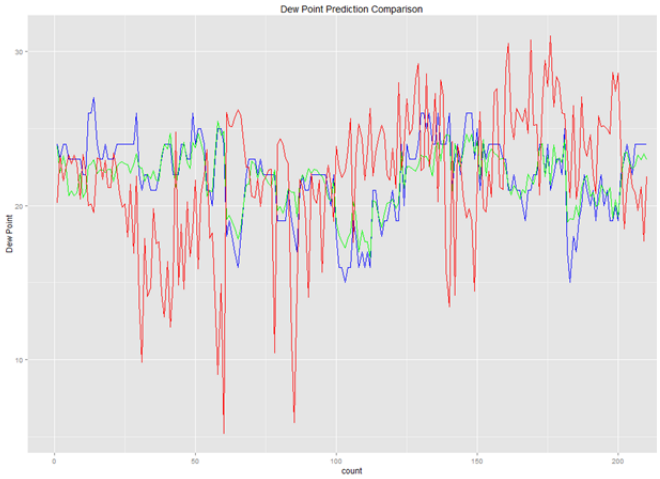

Fig. 6. Actual output, prediction by normal equation and prediction by gradient descent for dew-point

It is clear from Table V and Fig. 6 that the gradient descent method is unsuitable for predicting dew-point, whereas results produced by the normal equation method are highly reliable.

The graphs above proves the efficiency of the normal method. With larger dataset, better and more precision can be achieved.

V. CONCLUSION

This research suggests and proposes an efficient and accurate weather prediction and forecasting model using linear regression concepts and normal equation model. All these concepts are a part of machine learning. The normal equation is a very efficient weather prediction model and using the entities temperature, humidity and dew-point, it can be used to make reliable weather predictions. This model also facilitates decision making in day to day life. It can yield better results when applied to cleaner and larger datasets. Pre-processing of the datasets can be effective in the prediction as unprocessed data can also affect the efficiency of the model.

REFERENCES

[1] Radhika Y, Shashi M. *Atmospheric temperature prediction using support vector machines*. International Journal of Computer Theory and Engineering. 2009 Apr; 1(1):1793– 8201.

[2] Han J, Kamber *M. Data mining: Concepts and techniques*. Morgan and Kaufmann; 2000

[3] Smith BA, McClendon RW, Hoogenboom G. *Improving air temperature prediction with artificial neural networks*. International Journal of Computational Intelligence. 2007; 3(3):179–86.

[4] Hsieh WW, *Nonlinear canonical correlation analysis of the tropical pacific climate variability using a neural network approach.* Journal of Climate. 2001.

[5] Casas DM, Gonzalez AT, Rodrigue JEA, Pet JV. *Using data mining for short-term rainfall forecasting. Notes in Computer Science*. 2009; 5518:487–90.

[6] Smith BA, McClendon RW, Hoogenboom G. *Improving air temperature prediction with artificial neural networks. International Journal of Computational Intelligence*. 2007; 3(3):179–86.

[7] S. Sripada, E. Reiter, I. Davy, *SumTime-Mousam: Configurable marine weather forecast generator*, Expert (2003) 4–10.

[8] J.R. Quinlan, C4.5: *Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1992.

[9] P. Gorniak, D. Roy, *Grounded semantic composition for visual scenes*, J. Artificial Intelligence Res. 21 (2004) 429–470.