International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 158-162
Published Online October 2020 in IJEAST (http://www.ijeast.com)

WEATHER FORECASTING USING REGRESSION

Sahil Makhijani
Dept. Computer Engineering
Vidyalankar Polytechnic,
Mumbai, Maharashtra, India

Anurag Dubey
Dept. Computer Engineering
Vidyalankar Polytechnic
Mumbai, Maharashtra, India

Ankush Makhijani
Department of Science
Lakshya Prep College,
Mumbai, Maharashtra, India

Abstract— Weather forecasting has traditionally been done by complex simulation models of physics made up of complex equations of fluid dynamics and thermodynamics. Due to this, even minute errors in the simulation causes huge differences in the results. Thus, this method is inaccurate for long-term forecasting. On the other hand, machine learning techniques are more robust to minute errors in training datasets. In this paper we explore machine learning regression algorithm application to weather forecasting, to potentially generate more accurate weather forecasts for large periods of time. The regression model, described in this paper, will be able to capture trends in the weather. This model is outperformed by professional weather forecasting services.

Keywords*— Weather Forecasting, *Machine Learning, Regression

I. INTRODUCTION

Weather forecasting is a very important aspect of life. It is a product of science which has an influence on human life. Weather Forecasting is an essential area of analysis in everyday life. Weather for the future is one of the most important attributes to forecast because agriculture sectors, as well as many industries, are largely dependent on the weather conditions. Chonghua Yin et al.(2018) in his work mentioned that weather conditions are required to be predicted not only for future planning in agriculture and industries but also in many other fields like defense, mountaineering, shipping and aerospace navigation etc. It is often used to warn about natural disasters caused by abrupt change in climatic conditions.[1]

Mark Holmstrom, Dylan Liu and Christopher Vo et al. (2016) in their work elucidate that in early times, weather forecasting has been done by complex simulation models of physics made up of complex equations of fluid dynamics and thermodynamics. But what about the future?[2] As we say now that the future is AI. But, can AI protect the people from the destructive forces of nature? Well, the answer is yes, it can. The Age of AI Documentery expounds that one of the promises of AI is that it will enable us to use machine learning for prediction and conservation, anything from protecting wildlife to predicting Earthquakes. We know that AI cannot prevent disasters but we can agree that we need an equipment upgrade.[3] We use a regression algorithm which is one type of machine learning algorithm.

Weather forecasting is done using the data gathered by remote sensing satellites. Weather parameters like maximum temperature, minimum temperature, extent of rainfall, cloud conditions, wind streams and their directions, are projected using images and data taken by these meteorological satellites to access future trends. The variables defining weather conditions like temperature (maximum or minimum), relative humidity, rainfall etc., vary continuously with time, forming time series of each parameter and it is used to develop a forecasting model statistically that uses this time series data. The process of developing a forecast is explained in this paper.

II. REVIEW OF LITERATURE

When we hear the word machine learning we also hear a word, which is AI. There is always an ambiguity between machine learning and AI. AI is making computer smart which maximizes its chance of successfully achieving its goals. And Machine learning is AI using algorithms and statistical models. Machine learning has a great influence on human life. As quoted by a famous machine learning scientist, Andrew Ng, “It is difficult to think of a major industry that AI will not transform. This includes healthcare, education, transportation, retail, communications, and agriculture. There are surprisingly clear paths for AI to make a big difference in all of these industries.” And in this paper, we are focusing on the application of machine learning algorithm regression in the field of weather forecasting.

III. REGRESSION

Regression Analysis is a set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable’ or ‘target’) and one or more independent variables (often called 'features'). The most common form of regression analysis is linear regression, in which a researcher finds the line (or a more complex linear function) that most closely fits the data according to a specific mathematical criterion.

A. Simple Linear Regression

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 158-162
Published Online October 2020 in IJEAST (http://www.ijeast.com)

between x (input) and y (output). Hence, the name is Linear Regression. If we plot the independent variable (x) on the x-axis and dependent variable (y) on the y-axis, linear regression gives us a straight line that best fits the data points, as shown in the figure below.

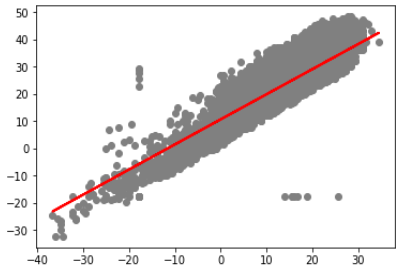

Fig 1: Simple linear regression graph

In the given example, we give minimum temperature (temperature during mid-night) of the day and we get maximum temperature (temperature during noon).

Equation of the line which best fits is:

$$h_\theta(x) = \theta x_i + b$$

Where,

$h_\theta(x)$ is hypothesis function
θ is weight
x_i is feature value
b is bias

Disadvantages of simple linear regression

1. Target depends on only one feature. (Weather depends on many features)
2. Linear Relationship between the feature and target. (Weather doesn't change linearly

B. Solution for 1: Multiple Linear Regression

Multiple linear regression (MLR) is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression (MLR) is to model the relationship between the explanatory and response variables.

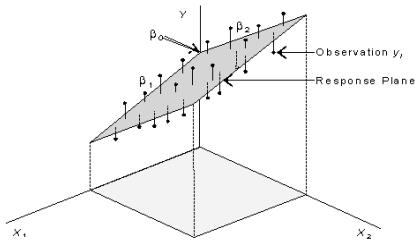

Fig 2: Multiple linear regression graph

For example, we can take the same graph of maximum temperature dependent of minimum temperature and we will also add one more feature of intensity of the sun.

Equation of plane which best fits is:

$$h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Where,

$h_\theta(x)$ is hypothesis function
θ_i are weights
x_i are features values
n is number of features

The selection of features is an important task. It directly affects the accuracy of our model. For example, salinity of the water, number of trees, topography, soil texture, distance from sea, sun intensity, etc. are an important feature in weather.

C. Solution for 2: Simple Polynomial Regression

What if your data is actually more complex than a simple straight line? Surprisingly, you can actually use a linear model to fit nonlinear data. A simple way to do this is to add powers of each feature as new features, then train a linear model on this extended set of features. This technique is called *Polynomial Regression.*

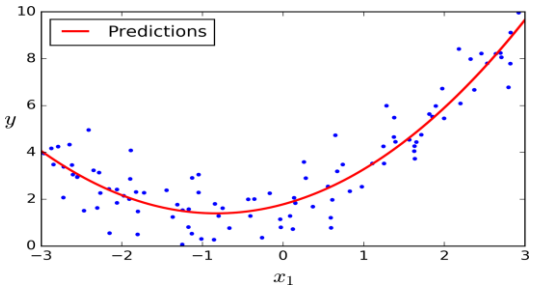

Fig 3: Simple Polynomial regression graph

For example, we can take the same maximum temperature and keep it with time as a feature. Now, the maximum temperature will decrease during winter and then increase during summer.

Equation of curve which best fits is:

$$h_\theta(x) = \theta_0 x^0 + \theta_1 x^1 + \theta_2 x^2 + \ldots + \theta_l x^l$$

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 158-162
Published Online October 2020 in IJEAST (http://www.ijeast.com)

Where,

$h_\theta(x)$ is hypothesis function
θ_i are weights
x is features values
l is degree of freedom

We have to choose a degree of freedom to have the best fit and avoid under fit as well as over fit.

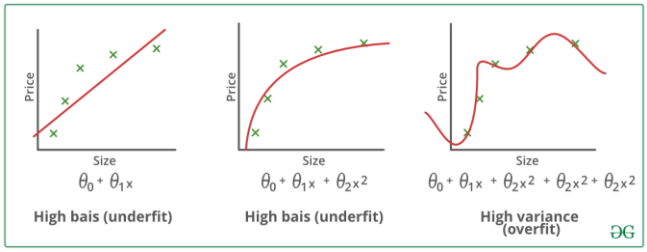

Fig 4: Underfit, Bestfit, Overfit

D. Combining both the solutions: Multiple Polynomial Regression

Here we combine both the solutions to get multiple features as well as non-linear relationship between features and target.

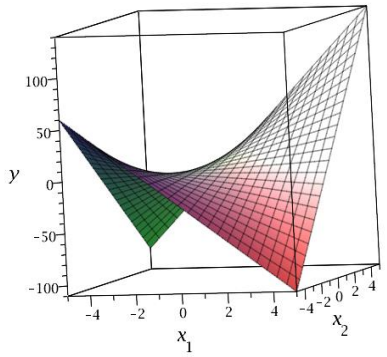

Fig 5: Multiple Polynomial regression graph

Maximum temperature depends upon many factors and it increases and decreases according to change in factor.

The equation of this shape is given by:

$$\begin{aligned} h_\theta(x) \;= \theta_0 x_0 &+ \theta_{11} x_1 + \theta_{12} x_2 + \ldots + \theta_{1n} x_n \\ &\theta_{21} x_1^2 + \theta_{22} x_2^2 + \ldots + \theta_{2n} x_n^2 \\ &\theta_{31} x_1^3 + \theta_{32} x_2^3 + \ldots + \theta_{3n} x_n^3 \\ &\vdots \\ &\theta_{l1} x_1^l + \theta_{l2} x_2^l + \ldots + \theta_{ln} x_n^l \end{aligned}$$

Where,

$h_\theta(x)$ is hypothesis function
θ_{ii} are weights
x_i are features values
n is number of features
l is degree of freedom

V. Cost Function

A cost function is a measure of how wrong the model is in terms of its ability to estimate the relationship between X and Y.

There are many cost functions. In this paper, we have used the mean squared error cost function.

We denote the cost function as J and it takes all the theta values.

$$J(\theta) \;=\; MSE \;=\; \frac{1}{n}\sum_{i=1}^{n} (h_\theta(x_i) - y_i)^2$$

Where,

$J(\theta)$ is Cost Function
MSE is Mean Squared Error Cost Function
n is number of data points of training data
$h_\theta(x)$ is hypothesis function
x , y are the data points

The objective of a ML model, therefore, is to find parameters, weights or a structure that minimizes the cost function. This is done by gradient descent

Vanilla Gradient Descent

Gradient Descent is a very generic optimization algorithm capable of finding optimal solutions to a wide range of problems. The general idea of Gradient Descent is to tweak parameters iteratively in order to minimize a cost function.

Process: Suppose you are lost in the mountains in a dense fog; you can only feel the slope of the ground below your feet. A good strategy to get to the bottom of the valley quickly is to go downhill in the direction of the steepest slope. This is exactly what Gradient Descent does: it measures the local gradient of the error function with regards to the parameter vector θ, and it goes in the direction of descending gradient. Once the gradient is zero, you have reached a minimum!

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 5, Issue 6, ISSN No. 2455-2143, Pages 158-162
Published Online October 2020 in IJEAST (http://www.ijeast.com)

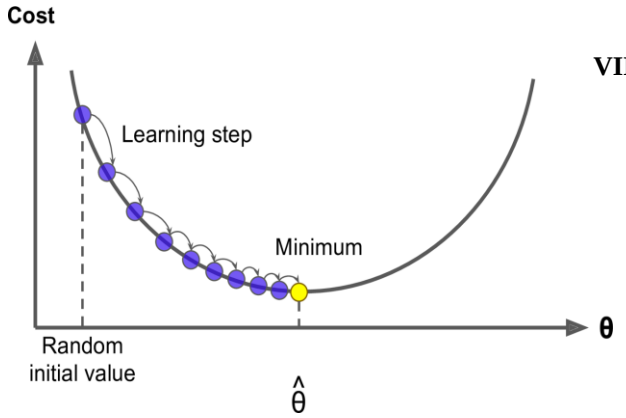

Fig 6: Gradient Descent Representation

To do this we iterate the following formula:

$$\theta_j := \theta_j - \alpha\left(\frac{\partial}{\partial\theta_j} J(\theta)\right)$$

Where,

θ_j is current value of θ

:= is symbol of changing value in iteration (Not Mathematical Equals To)

α is factor of taking steps

$J(\theta)$ is Cost Function

Thus, we get values of theta which will make the shape best fit the data.

VI. Training and Testing Datasets

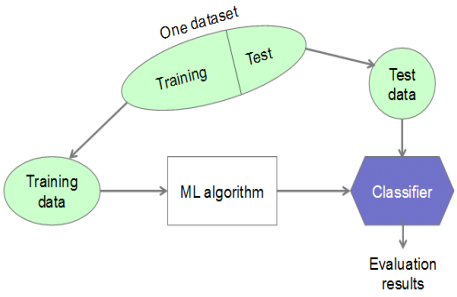

Fig 7: Dividing dataset into training and testing

We divide our dataset into training and testing dataset.

- Training dataset (usually 80%) is used to train the machine learning model and to get the best fit shape.
- Testing dataset (usually 20%) is used to test the model formed.

This helps in knowing the accuracy of the model.

Also, we get to know whether the degree of freedom we took is making over fit or under fit or best fit.

VII. Process of Weather Prediction

1. Data Collection
2. Feature Selection
3. Dividing dataset into testing and training datasets
4. Train the model with Training dataset.
    a. Create Hypothesis Function
    b. Create Cost Function
    c. Apply Gradient Descent
5. Test the model with Testing dataset
6. Repeat the process with different degree of freedom if it under fits or over fits
7. Get the highest accuracy of the model
8. Predict the future weather by giving all the values of feature expected in future

VIII. Conclusion

In this paper, we presented a technology to utilize machine learning techniques to provide weather forecasts. Machine learning technology can provide intelligent models, which are much simpler than traditional physical models. They are less resource-hungry and can easily be run on almost any computer including mobile devices. Our evaluation results show that these machine learning models can predict weather features accurately enough to compete with traditional models. We also utilize the historical data from surrounding areas to predict the weather of a particular area. We show that it is more effective than considering only the area for which weather forecasting is done. We can use weather forecasts for saving lives too. For example, we can make arrangements for the local people in the area where famine is going to occur. AI might not prevent disasters but new scientific tools like machine learning, image recognition and productive modelling might help us get ahead of them.

In future, we have plans to utilize low-cost Internet of Things (IoT) devices, such as temperature and humidity sensors, in collecting weather data from different parts of a city. The use of different sensors could increase the number of local features in the training dataset. This data, along with the weather station data, will further improve the performance of our prediction models.

IX. Acknoledgements

1) Thanks to Dr. Pooja Srivastava for guidance and motivation.
2) Thanks to our parents for continuous support.

X. REFERENCES

[1] Chonghua Yin. (2018). Regression Analysis for Weather Forecasting.

https://www.linkedin.com/pulse/regression-analysis-weather-forecasting-chonghua-yin

[2] Mark Holmstrom, Dylan Liu and Christopher Vo. (2016). Machine Learning Applied to Weather Forecasting.

http://cs229.stanford.edu/proj2016/report/HolmstromLiuVo-MachineLearningAppliedToWeatherForecasting-report.pdf

[3] YouTube Originals. (2016). The Age of AI. S1 E7.

https://www.youtube.com/watch?v=0wy4u34fii4&vl=en

[4] Aurélien Géron. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition.