Weather Prediction

Anirudh Kumar Singh

Anirudh kumar Singh
16SCSE107002
1613107008

1. Introduction:

1.1 Description:

General circulation models work as tools for weather forecasting. There exists lots of assumption based analysis on weather conditions presented us with the best prediction models like Numerical weather prediction, climatic variability assessment, weather Surveillances radar, predictions of Warming, etc.

The Scientists are still in working process of overcoming the limitations of computer models to improvise the accuracy rate of prediction through recent technologies of adding intelligence to machine. To add intelligence the system as human we have given a study platform called Artificial Neural networks, Machine learning, rule based techniques where there exist ample impetus to study the weather occurrence and prediction.

Weather prediction is a convenient case for studying machine learning. By developing APIs for accessing available data from meteorological institutes and other weather stations, this gives access to an abundance of data. Weather data is something that most people can relate to in their daily life, but is also important for energy systems, flood prediction, etc. Good physical based meteorological models are available, which makes it easy to compare the quality of machine learning models. Large Scale weather prediction information can easily be accessible from weather report services for a big region or a city. They report like tomorrow heavy rain expected in this city and visibility will be 10 km. and so on. However, large scale static forecasts are not sufficiently stable on a small scale where local effects and regency become significant or predominant.

1

1.2 Purpose:

Every Human subject to adjust themselves with respect to weather conditions for their dressing habits to strategic organizational planning activities, since the adverse weather conditions may cause a considerable damage to lives and properties.

We need to be on alert for these adverse weather conditions by taking some precautions and using prediction mechanisms to detect them and provide early warning of hazardous weather phenomena. Weather prediction is an indispensable requirement for all of us.

Weather is important for most aspects of human life. Predicting weather is very useful. Humans have attempted to make predictions about the weather, many early religions used gods to explain the weather. Only relatively recently have humans developed reasonably accurate weather predictions. We decided to collect weather data and measured the accuracy of predictions made using linear regression.

The Weather prediction model designed by us would be of great use to the farmers and for normal being as well. This model basically uses historical weather data to predict the weather on a specific day of and year in the future. Initially the aim is to teach the model with large historical data set and then use it for weather prediction.

The observations include:

- **Temperature** - the measure of warmth or coldness
- **Humidity** - the amount of moisture in the atmosphere
- **Precipitation** - the amount of moisture (usually rain or snow) which falls on the ground
- **Wind Speed** - the speed at which air flows through the environment
- **Wind Direction** – the direction in which the wind is moving
- **Pressure** - the force the atmosphere applies on the environment

2

1.3 Scope:

There is a general and increasing interest on weather information, since every day we habitually give an ear to weather forecast news for local and large-scale long-term or short-term weather predictions. Leading weather research institutions and companies have been developing weather prediction systems capable of detecting, predicting and forecasting weather phenomena and hazards by utilizing state-of-the-science technologies. Thus weather prediction utilization fields and prediction accuracy increases monotonically by the time.

3

2. Literature Survey:

Machine Learning is the study of computer algorithms that improve automatically through experience. Applications range from data mining programs that discover general rules in large data sets, to information filtering systems that automatically learn users' interests.

Machine Learning is concerned with the design and development of algorithms that allow computers to evolve behaviors based on empirical data, such as from sensor data or databases. A major focus of Machine Learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data; the difficulty lies in the fact that the set of all possible behaviors given all possible inputs is too complex to describe generally in programming languages, so that in effect programs must automatically describe programs.

In recent years many successful machine learning applications have been developed, ranging from data-mining programs that learn to detect fraudulent credit card transactions, to information filtering systems that learn users' reading preferences, to autonomous vehicles that learn to drive .On public highways. At the same time, there have been important advances in the theory and algorithms that form the foundations of this field.

The poor performance results produced by statistical estimation models have flooded the estimation area for over the last decade. Their inability to handle categorical data, cope with missing data points, spread of data points and most importantly lack of reasoning capabilities has triggered an increase in the number of studies using non-traditional methods like machine learning techniques. The area of machine learning draws on concepts from diverse fields such as statistics, artificial intelligence, philosophy,

Anirudh kumar Singh
16SCSE107002
1613107008

4

information theory, biology, cognitive science, computational complexity and control theory.

There are two main types of Machine Learning algorithms. In this project, supervised learning is adopted here to build models from raw data and perform regression and classification.

Supervised learning: Supervised Learning is a machine learning paradigm for acquiring the input output relationship information of a system based on a given set of paired input-output training samples. As the output is regarded as the label of the input data or the supervision, an input-output training sample is also called labeled training data, or supervised data. Learning from Labeled Data, or Inductive Machine Learning.

The goal of supervised learning is to build an artificial system that can learn the mapping between the input and the output, and can predict the output of the system given new inputs. If the output takes a finite set of discrete values that indicate the class labels of the input, the learned mapping leads to the classification of the input data. If the output takes continuous values, it leads to a regression of the input. It deduces a function from training data that maps inputs to the expected outcomes. The output of the function can be a predicted continuous value (called regression), or a predicted class label from a discrete set for the input object (called classification). The goal of the supervised learner is to predict the value of the function for any valid input object from a number of training examples. The most widely used classifiers are the Neural Network (Multilayer perceptron), Support Vector Machines, k-nearest neighbour algorithm, Regression Analysis, Artificial neural networks and time series analysis.

Unsupervised learning: Unsupervised learning studies how systems can learn to represent

particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. By contrast with supervised learning or reinforcement learning, there are no explicit target outputs or environmental evaluations associated with each input; rather the unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output.

Anirudh kumar Singh
16SCSE107002
1613107008

6

3. Proposed model:

Our model consists of collecting the historical weather data that includes various important factors responsible for the weather change that includes the temperature, both the maximum and minimum temperature, the moisture or humidity in the atmosphere, precipitation, UV Index of the atmosphere and the mean pressure of the atmosphere.

In our proposed model the collected dataset is segregated into the parts which are of use and which aren't of any use to the machine learning model. After that the dataset goes through the data preprocessing part wherein the data is passed on to a process where the missing and the error values in the dataset are replaced by the mean values or the most occurring value in that filed. Other way is not to consider those values and replacing those empty values with EAN and the carrying out the other tasks.

After the data preprocessing is completed there comes the part wherein the cleaned dataset is segregated into two parts namely the training set and the test set. The training set is used to train the machine learning model to teach the model to compute the results and the testing set is the used to find the results and then comparing the actual and the calculated value and the using the error value as the benchmark to teach the machine learning model further.

The training phase also will also contain fold cross validation wherein the dataset is divided in to k sets k times and then the dataset is divided into test and training sets such that in a set training sets are selected randomly and then in sets the model is trained. The kth set is then used as the test set for the trained machine learning model.

This technique not only does helps us to reduce the condition of underfitting and the condition of overfitting aswell.

7

Anirudh kumar Singh
16SCSE107002
1613107008

3.1 Algorithm:

---

1. Train model with x where x ∈ W. W={ meanT, meanH, meanP,WindSpeed} of past few years.

2. Clean the dataset by removing or replacing the missing values.

3. Evaluate and optimize the model with test set.

4. a. Remove the outliers.
   b. Minimize the cost function.

5. Input a range of days (after the model is ready).

6. Predict the meanT for the desired day or a range of days be it for the future or in the past.

3.2 Module split-up:

- Data Set
- Data Preprocessing
- Splitting into Train and Test set

Anirudh kumar Singh
16SCSE107002
1613107008

- Extracting and Examining outcome
- Deriving relationships
- Plotting graph
- Deriving best match line

This method is mostly used for forecasting and finding out cause and effect relationship between variables. Regression techniques mostly differ based on the number of independent variables and the type of relationship between the independent and dependent variables.

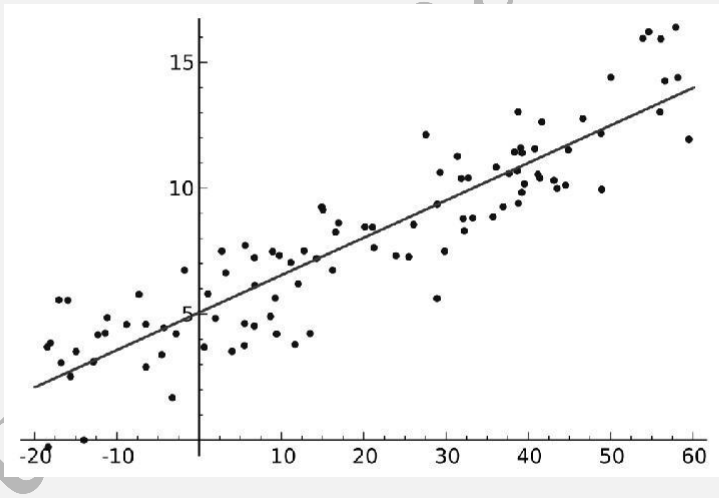

Fig.1 Sample Best fit line

linear regression is a type of regression analysis where the number of independent variables is one and there is a linear relationship between the independent(x) and dependent(y) variable. The red line in the above graph is referred to as the best fit straight line. Based on the given data points, we try to plot a line that models the points the best. The line can be modelled based on the linear equation shown below.

y = a_0 + a_1 * x   -- Linear Equation

The motive of the linear regression algorithm is to find the best values for a_0 and a_1. Before moving on to the algorithm, let's have a look at two important concepts you must know to better understand linear regression.

3.3 Cost Function:

The cost function helps us to figure out the best possible values for a_0 and a_1 which would provide the best fit line for the data points. Since we want the best values for a_0 and a_1, we convert this search problem into a minimization problem where we would like to minimize the error between the predicted value and the actual value.

$$minimize \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

$$J = \frac{1}{n} \sum_{i=1}^{n} (pred_i - y_i)^2$$

3.3.1 Minimization and Cost Function:

We choose the above function to minimize. The difference between the predicted values and ground truth measures the error difference. We square the error difference and sum over all data points and divide that value by the total number of data points. This provides the average squared error over all the data points. Therefore, this cost function is also known as the Mean Squared Error(MSE) function. Now, using this MSE function we are going to change the values of a0 and a1 such that the MSE value settles at the minima.

3.4 Gradient Descent:

The next important concept needed to understand linear regression is gradient descent. Gradient descent is a method of updating a0 and a1 to reduce the cost function(MSE). The idea is that we start with some values for a0 and a1 and then we change these values iteratively to reduce the cost. Gradient descent helps us on how to change the values.

Anirudh kumar Singh
16SCSE107002
1613107008

10

It is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our model. Parameters refer to coefficients in Linear Regression and weights in neural networks.

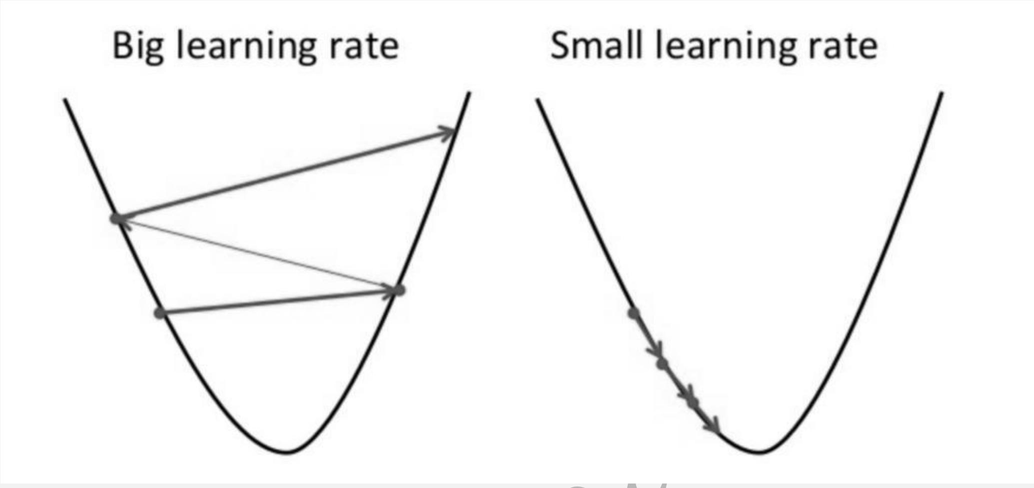

Fig.2 Gradient Descent

To draw an analogy, imagine a pit in the shape of U and you are standing at the topmost point in the pit and your objective is to reach the bottom of the pit. There is a catch, you can only take a discrete number of steps to reach the bottom. If you decide to take one step at a time you would eventually reach the bottom of the pit but this would take a longer time. If you choose to take longer steps each time, you would reach sooner but, there is a chance that you could overshoot the bottom of the pit and not exactly at the bottom. In the gradient descent algorithm, the number of steps you take is the learning rate. This decides on how fast the algorithm converges to the minima.

3.4.1 Learning rate:

The size of these steps is called the learning rate. With a high learning rate we can cover more ground each step, but we risk overshooting the lowest point since the slope of the hill is constantly changing. With a very low learning rate, we can confidently move in the

11

direction of the negative gradient since we are recalculating it so frequently. A low learning rate is more precise, but calculating the gradient is time-consuming, so it will take us a very long time to get to the bottom.

3.5 Feature scaling :

It is a method used to standardize the range of independent variables or features of data. In data processing, it is also known as data normalization and is generally performed during the data preprocessing step.

3.5.1 Where to Apply Feature Scaling:

Real world dataset contains features that highly vary in magnitudes, units, and range. Normalisation should be performed when the scale of a feature is irrelevant or misleading and not should Normalise when the the scale is meaningful.

The algorithms which use Euclidean Distance measure are sensitive to Magnitudes. Here feature scaling helps to weigh all the features equally.

Formally, If a feature in the dataset is big in scale compared to others then in algorithms where Euclidean distance is measured this big scaled feature becomes dominating and needs to be normalized.

12

3.5.2Methods:

3.5.2.1 Rescaling (min-max normalization):

Also known as min-max scaling or min-max normalization, is the simplest method and consists in rescaling the range of features to scale the range in [0, 1] or [−1, 1]. Selecting the target range depends on the nature of the data. The general formula is given as:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where x is an original value, x' is the normalized value.

3.5.2.2 Mean normalization:

$$x' = \frac{x - \text{average}(x)}{\max(x) - \min(x)}$$

where x is an original value, x' is the normalized value.

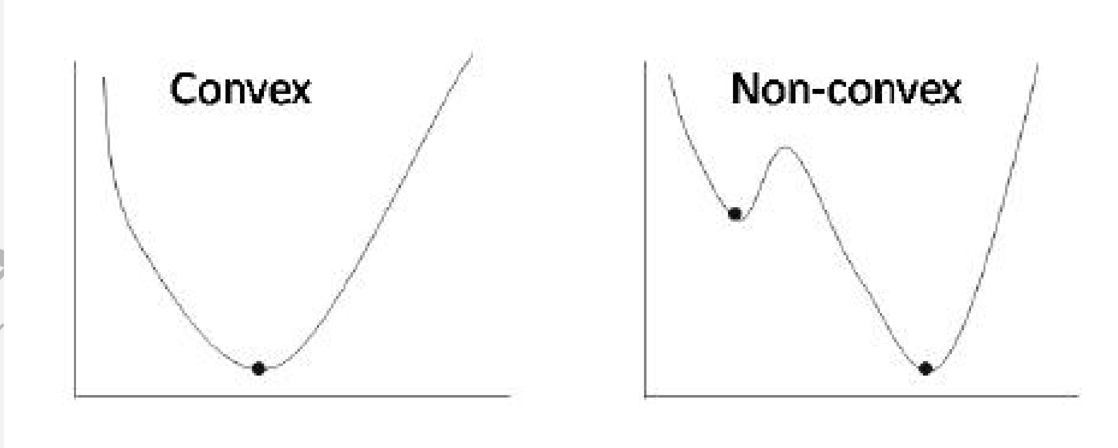

Fig.3 Convex vs Non-convex function

Anirudh kumar Singh
16SCSE107002
1613107008

13

Anirudh kumar Singh
16SCSE107002
1613107008

Sometimes the cost function can be a non-convex function where you could settle at a local minima but for linear regression, it is always a convex function.

You may be wondering how to use gradient descent to update a0 and a1. To update a0 and a1, we take gradients from the cost function. To find these gradients, we take partial derivatives with respect to a0 and a1. Now, to understand how the partial derivatives are found below you would require some calculus but if you don't, it is alright. You can take it as it is.

$$J = \frac{1}{n}\sum_{i=1}^{n}(pred_i - y_i)^2$$

$$J = \frac{1}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i)^2$$

$$\frac{\partial J}{\partial a_0} = \frac{2}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i) \implies \frac{\partial J}{\partial a_0} = \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i)$$

$$\frac{\partial J}{\partial a_1} = \frac{2}{n}\sum_{i=1}^{n}(a_0 + a_1 \cdot x_i - y_i) \cdot x_i \implies \frac{\partial J}{\partial a_1} = \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i) \cdot x_i$$

$$a_0 = a_0 - \alpha \cdot \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i)$$

$$a_1 = a_1 - \alpha \cdot \frac{2}{n}\sum_{i=1}^{n}(pred_i - y_i) \cdot x_i$$

14

The partial derivates are the gradients and they are used to update the values of a0 and a1. Alpha is the learning rate which is a hyperparameter that you must specify. A smaller learning rate could get you closer to the minima but takes more time to reach the minima, a larger learning rate converges sooner but there is a chance that you could overshoot the minima.

The cost function that linear regression seeks to minimize is:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m} \|h_\theta(x^{(i)}) - y^{(i)}\|^2, . \qquad (1)$$

where m is the number of training examples. The value of that minimizes the cost in equation 1 is:

$$\theta = (X^T X)^{-1} X^T Y. \qquad (2)$$

After cleaning of the data and applying the required data preprocessing techniques we plot the graph to find if the features we are using are correlated or not.

A model with high descriptive power is prone to overfitting. The term over-fitting is used in empirical modeling to describe what happens when a model adapts to random variations in the training set which does not generalize well to new data. This effect is apparent in all forms of empirical modeling, from simple curve fitting to complex ML models.

Since the ML models have such a large number of coefficients, the over-fitting problem is particularly important. The simplest way to reduce the risk of overfitting is to increase

the amount of training data, either by collecting more data or artificially creating more training data through some form of transformation on the original data.

Another way to reduce the risk of over-fitting is to apply a regularization method. The subject of regularization is a research field in itself, which involves methods that prevents the training algorithms for ML models from adapting to random variations in the training data.

The simplest form of regularization is to have a large amount of data. Since this work is based on retrieving weather data from online databases, the cost of obtaining data is relatively low, hence a large amount of training data is readily available.

3.6 Mean Square Error:

In statistics, the mean squared error (MSE) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors that is, the average squared difference between the estimated values and what is estimated. MSE is a risk function, corresponding to the expected value of the squared error loss. The fact that MSE is almost always strictly positive (and not zero) is because of randomness or because the estimator does not account for information that could produce a more accurate estimate.

$$\mathrm{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

In regression analysis, the term mean squared error is sometimes used to refer to the unbiased estimate of error variance: the residual sum of squares divided by the number of degrees of freedom. This definition for a known, computed quantity differs from the above definition for the computed MSE of a predictor in that a different denominator is used. The denominator is the sample size reduced by the number of model parameters estimated from the same data, (n-p) for p regressors or (n-p-1) if an intercept is used.

16

Also in regression analysis, mean squared error, often referred to as mean squared prediction error or out-of-sample mean squared error, can refer to the mean value of the squared deviations of the predictions from the true values, over an out-of-sample test space, generated by a model estimated over a particular sample space. This also is a known, computed quantity, and it varies by sample and by out-of-sample test space.

An MSE of zero, meaning that the estimator x predicts observations of the parameter x with perfect accuracy, is the ideal, but is typically not possible.

Both linear regression techniques such as analysis of variance estimate the MSE as part of the analysis and use the estimated MSE to determine the statistical significance of the factors or predictors under study. The goal of experimental design is to construct experiments in such a way that when the observations are analyzed, the MSE is close to zero relative to the magnitude of at least one of the estimated treatment effects.

MSE is also used in several stepwise regression techniques as part of the determination as to how many predictors from a candidate set to include in a model for a given set of observations.

The error depends on the scale of the data. It cannot be compared across different data, and there is no "default" of what is a good value. It depends on what an acceptable prognosis error is for you.

MSE basically measures average squared error of our predictions. For each point, it calculates square difference between the predictions and the target and then average those values.

The higher this value, the worse the model is. It is never negative, since we’re squaring the individual prediction-wise errors before summing them, but would be zero for a perfect model .

3.6.1 Advantage:

Useful if we have unexpected values that we should care about. Very high or low value that we should pay attention.

3.6.2 Disadvantage:

If we make a single very bad prediction, the squaring will make the error even worse and it may skew the metric towards overestimating the model's badness. That is a particularly problematic behaviour if we have noisy data (that is, data that for whatever reason is not entirely reliable)—even a ―perfect‖ model may have a high MSE in that situation, so it becomes hard to judge how well the model is performing. On the other hand, if all the errors are small, or rather, smaller than 1, than the opposite effect is felt: we may underestimate the model's badness.

3.7 Flow Chart:

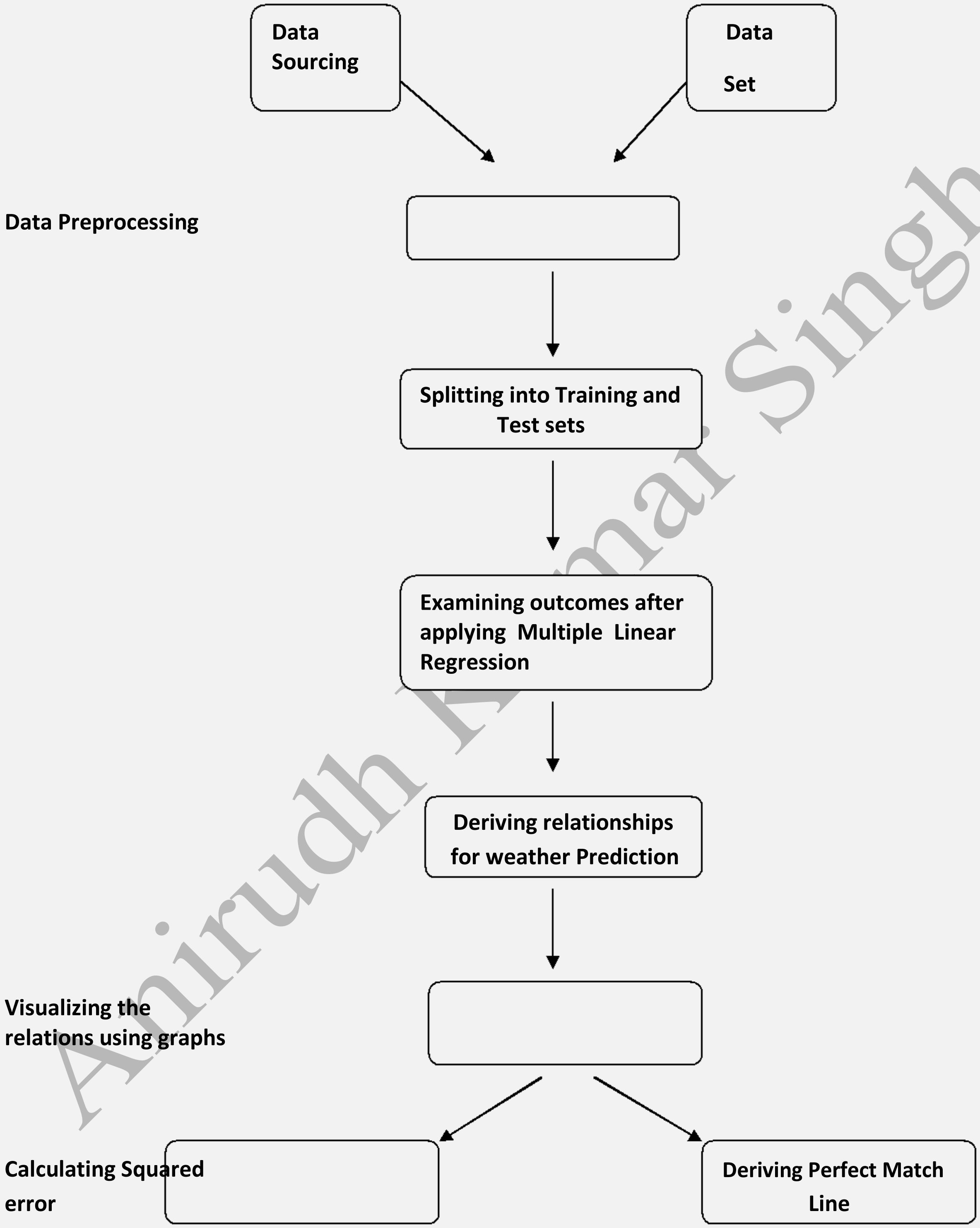

Fig. 4 Flow Chart

4. Implementation:

4.1 The Dataset:

- The Data Set we are using has been acquired by a website named https://rp5.ru/

- It provides Present as well as Historical weather data several years back in time depending upon the location of the weather stations.

- The data set provided by the source is kind of reliable and has ample features to work on example mean temp, max temp, humidity, precipitation etc.

- The data set we are using is of a specific area(New Delhi) and is categorised date wise.

4.2 Tools Used:

- **Spyder IDE** : Powerful scientific environment written in python for the data analysts and engineers.

- **GNU Octave** : Powerful mathematics-oriented syntax with built-in plotting and visualization tools.

- **Jupyter Notebook** : Open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text.

20

4.3 Libraries used:

- **Sklearn** : Machine Learning library that helps in making ML models.
- **Matplotlib** : Used for plotting intuitive charts.
- **Pandas** : Import and manage datasets.
- **Numpy** : A mathematical Library.
- **Future** : getting accustomed to incompatible changes that are used in the future releases.

4.4 Selecting Features for our Model :

A key assumption required by the linear regression technique is that we have a linear relationship between the dependent variable and each independent variable. One way to assess the linearity between our independent variable, which for now will be the mean temperature, and the other independent variables is to calculate the Pearson correlation coefficient.

The Pearson correlation coefficient is a measurement of the amount of linear correlation between equal length arrays which outputs a value ranging -1 to 1. Correlation values ranging from 0 to 1 represent increasingly strong positive correlation. By this I mean that two data series are positively correlated when values in one data series increase simultaneously with the values in the other series and, as they both go up in increasingly equal magnitude the Pearson correlation value will approach 1.

Correlation values from 0 to -1 are said to be inversely, or negatively, correlated in that when the values of one series increase the corresponding values in the opposite series

decrease but, as changes in magnitude between the series become equal (with opposite direction) the correlation value will approach -1. Pearson correlation values that closely straddle either side of zero are suggestive to have a weak linear relationship, becoming weaker as the value approaches zero.

Below are the correlation values :0.8-1.0 - Very Strong

0.6-0.8 - Strong

0.4-0.6 -Moderate

0.2-0.4 - Weak

0.0-0.2 – Very Weak

4.5 Visualizing the Relationships :

For this plot I would like to have the dependent variable "hourlymeantemp" be the consistent y-axis along all of the predictor variables plots. One way to accomplish this is to create a grid of plots. Pandas This can be done using Octave or Pandas function called Scatter_plot().

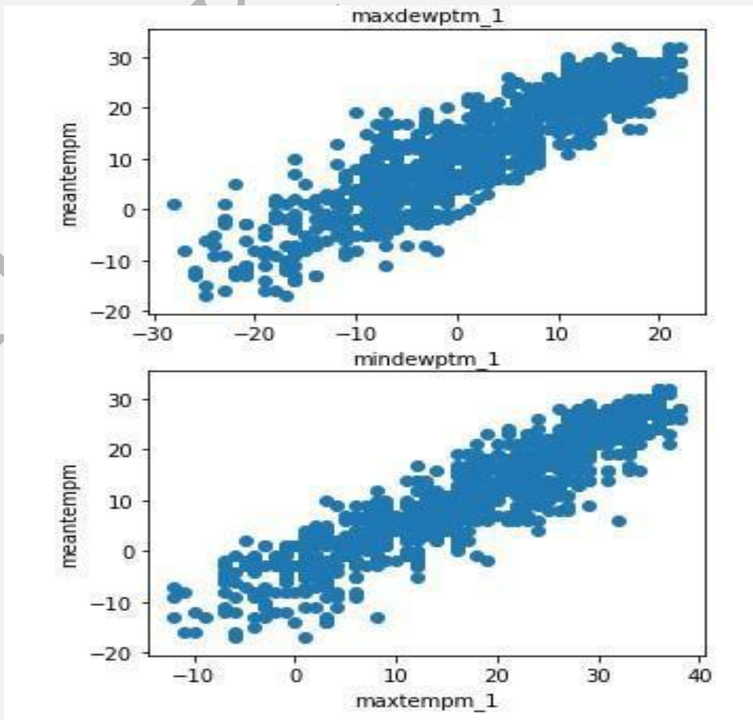

Fig. 5 Correlation

Anirudh kumar Singh
16SCSE107002
1613107008

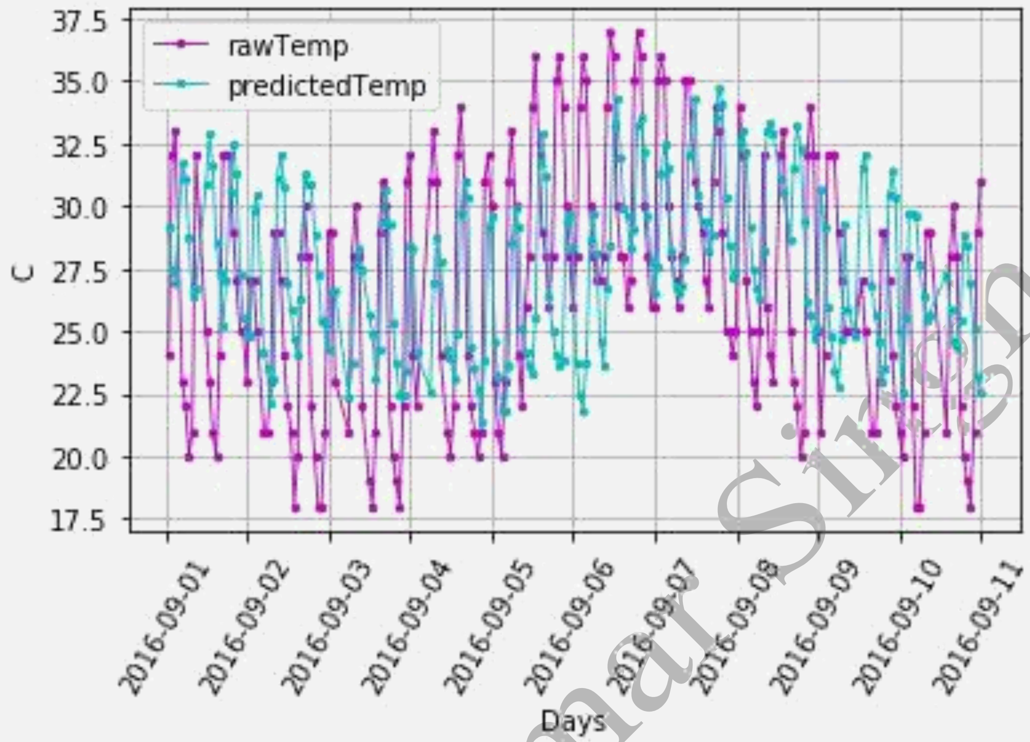

Fig. 6 Actual Temp vs Predicted Temp.

- The displayed graph is a comparison between the actual temperature and the predicted temperature over that range.

- As depicted the X-axis bears the range of days and Y-axis bears the temperature in C.

- At several places the ML model fails to predict the temperature about the actual temperature range, this can be due to various reasons be it the number of features, the variance of the model or maybe due to anomaly in the dataset.

23

4.5.1 Best Fit Line:

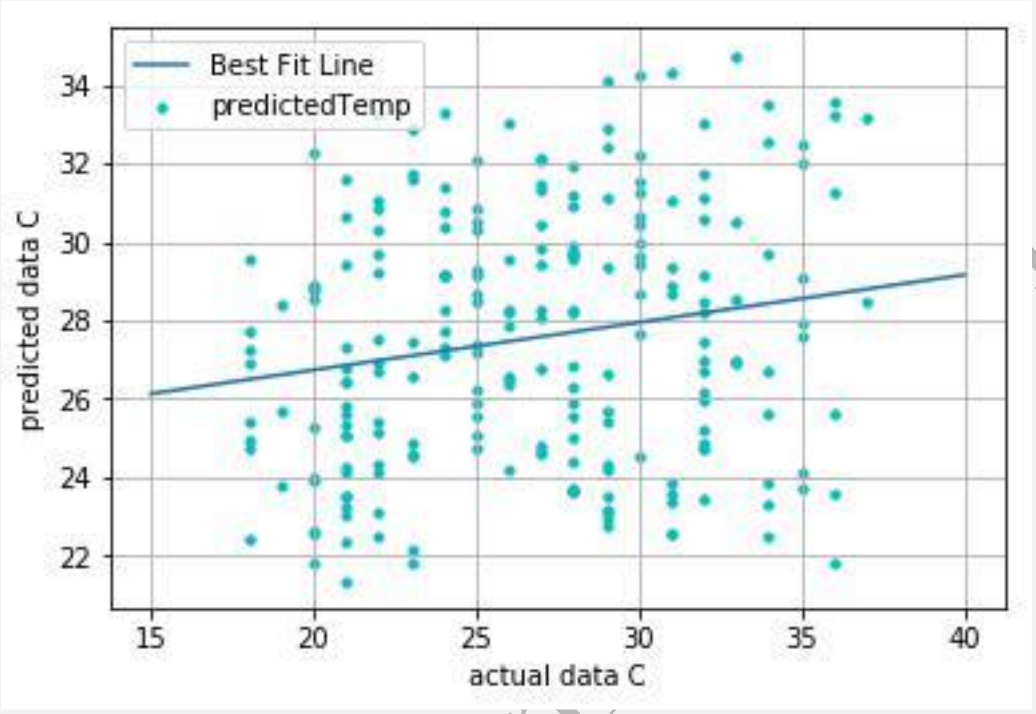

Fig. 7 Best fit Line

- The graph depicts the best match line for the predicted and actual data.

- It does contains some outliers and this is because the outliers have not been removed.

- The depicted graph bears the same range of time.

- For Instance, ten day range of a particular month is depicted in the graph.

24

4.6 Actual Screenshots:

```
In [16]: runfile('C:/Users/YS/Desktop/fyp/fyp3.py', wdir='C:/Users/YS/Desktop/fyp')
C:/Users/YS/Desktop/fyp/fyp3.py:1: DtypeWarning: Columns (15) have mixed types.
Specify dtype option on import or set low_memory=False.
  from __future__ import division
D:\Programs\Anaconda3\lib\site-packages\pandas\core\indexing.py:537:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/
indexing.html#indexing-view-versus-copy
  self.obj[item] = s
D:\Programs\Anaconda3\lib\site-packages\pandas\core\frame.py:3027:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/
indexing.html#indexing-view-versus-copy
  return super(DataFrame, self).rename(**kwargs)
D:\Programs\Anaconda3\lib\site-packages\pandas\core\indexing.py:194:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/
indexing.html#indexing-view-versus-copy
  self._setitem_with_indexer(indexer, value)
C:/Users/YS/Desktop/fyp/fyp3.py:364: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/
indexing.html#indexing-view-versus-copy
  df_time_train.loc[:, new_target] =
df_time_train[new_target].interpolate(method='time')
```

Fig. 8 Screenshot 1

Anirudh kumar Singh
16SCSE107002
1613107008

```
C:/Users/YS/Desktop/fyp/fyp3.py:40: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/
indexing.html#indexing-view-versus-copy
  df_re.loc[:, features] = df[features].interpolate(method='time', inplace = False)
Mean squared error for Train set: 36.575

Mean squared error for Test Set: 38.349

Variance score (test): 0.491

Plotting....
Prediction start from 2016-09-01 00:00:00

Prediction end at 2016-09-11 00:00:00
```

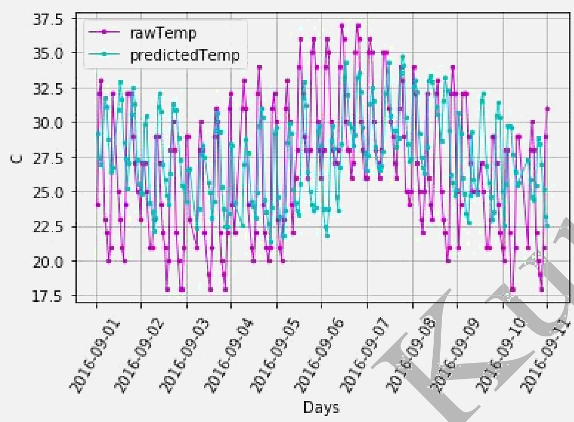

Fig. 9 Screenshot 2

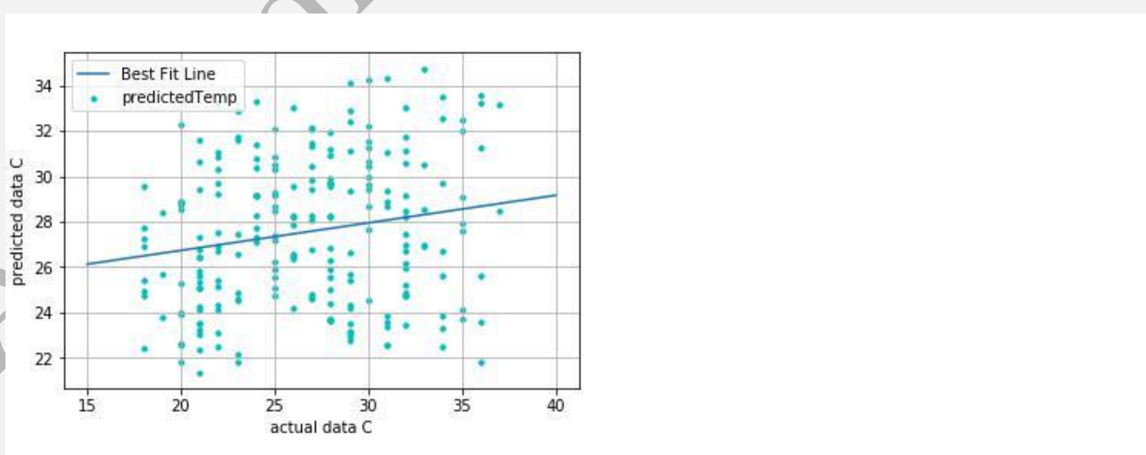

Fig. 10 Screenshot 3

26

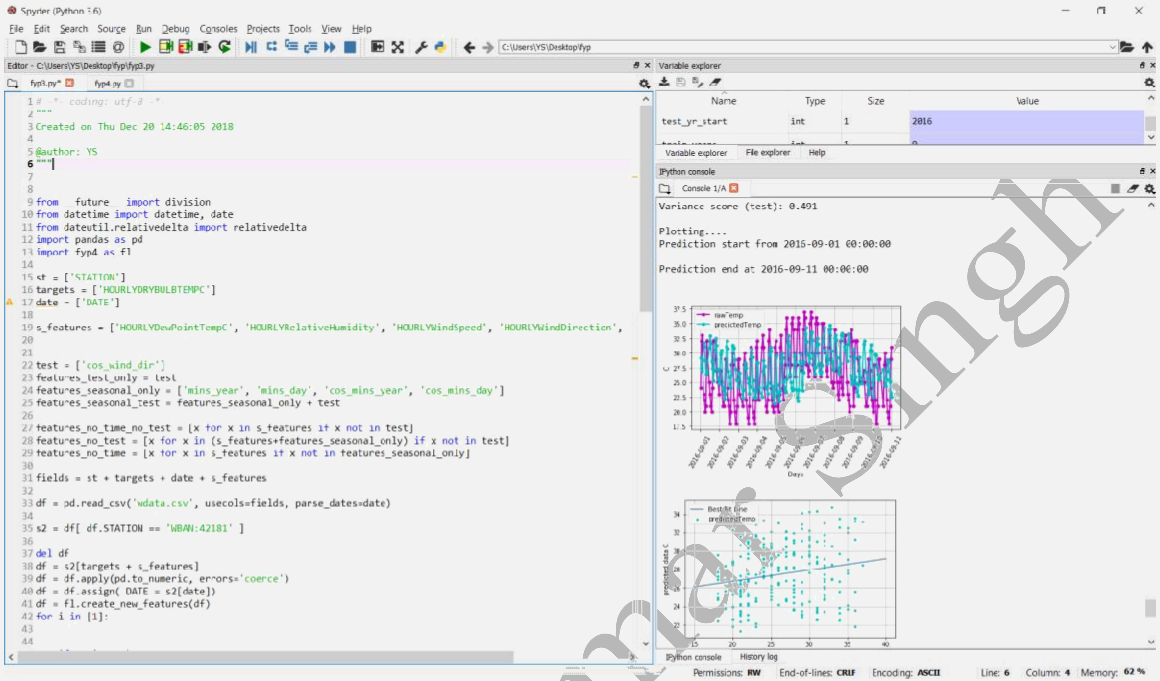

Fig. 11 Screenshot 4

4.7 Website Implementation

The above work and the results have been given an outline using a website. All the necessary information regarding this project is present on the website that i have build.

The said website has the following sections:

- Home
- Introduction
- Dataset
- Tools Used
- Result
- Conclusion
- About

Moreover the dataset used in this project is also available on the website mentioned below.

The link for the website is given below:

https://anirudhsingh0703.wixsite.com/weatherprediction

5.Results and Discussions:

- In the ML model used in this project we found that the quadratic hypothesis cannot be used for this purpose. This is because a quadratic equation comes down and we do not want that to happen.

- Cubic or higher degree hypothesis function should be used. Moreover feature scaling plays an important role as the value is increasing as per the degree of the polynomial.

6.Conclusions and Future Works:

- During this project we found that feature scaling is an important aspect of ML models. The basic idea is to make sure that the features are on a similar scale.

- Here we are only trying to speed up the things, the goal is to get all the input variables into roughly one of these ranges, give or take a few.

- For the upcoming years we should try to minimize the variance as far as possible which could help yield better prediction thereby resulting in a successful ML model.

- Since the outliers are not good for a ML model they should be avoided or removed before finding out the best fit, this not only does increases the accuracy of the model it also maintains the consistency of the results which may otherwise be differed when outliers are included.

- Professional weather forecasters are not perfect, but their predictions are typically more accurate than those of this linear regression model. This implies that weather is a non-linear system. Additionally, my predictions were all based on data from one location as opposed to multiple locations that most forecasters use. Though my model is imperfect, it does describe limitation of linear regression on predicting weather.

Anirudh kumar Singh
16SCSE107002
1613107008

Anirudh kumar Singh
16SCSE107002
1613107008