News Posts' Shares predicted through Regression and Classification

Under the guidance of
Prof R Jha

Presented by
Divyanshu Shekhar
(BE/10253/2013)

Rise of Machine Learning

- Grew out of work in AI
- New capability of Computers

Wide Applications of ML

- Database Mining
  - Large datasets from the growth of automation/web.
  - E.g web click data, medical records, engineering
- Applications can’t program by hand
  - E.g. Autonomous Helicopter, Handwriting Recognition, Natural Language Processing, Computer Vision.
- Self Customizing Programs
  - E.g. Amazon, Google Recommendations
- Understanding Human Learning (Brain, Real AI)
- Large scale companies e.g. Google, facebook etc are tremendously investing in this modern area.

What is Machine Learning?

- “ML is the field of study that gives computers the ability to learn without being explicitly programmed” - Arthur Samuel.
- “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P, improves with experience E” - Tom Mitchell.
- AI Dream- To build machines as intelligent as human. The best solution is to learn algorithms trying to mimic how the human brain learns.

Classification of ML problems

- Supervised Learning
  - In supervised learning, we are given a data set and already know what our correct output should look like, having the idea that there is a relationship between the input and the output.

- Unsupervised Learning
  - Concern is to find a structure in the dataset where we don't necessarily know the effects of the variables and possess little or no Idea what our results should look like.
  - e.g. Social Media, Market Segmentation

Supervised Learning

- Regression
  - results within a continuous output
  - e.g - given a picture of person, predict the age of the person.
- Classification
  - Discrete Output
  - e.g. - Given a patient with tumor, predict whether the tumor is benign or malignant.

Literature Review I

- A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance
- This paper's conclusion tells that classification algorithm produces a better performance than the regression algorithms.
- This helps in solidifying the concept that prediction must be based on several techniques and the analysis of those results obtained.

Literature Review II

- A Comparison Of Logistic Regression, Neural Networks, and Classification trees in a study to predict success of actuarial students.
- The conclusion of this papers says that although logistic regression method works well for a variety of problems but more accurate results can be obtained by deploying different other algorithms

Selection of Dataset

While selecting the dataset, few points were kept in mind

- Selecting a dataset which can be utilised for regression and classification problem as well.
- Have training examples in the order of 10, 000
- Multivariable features

Online News Popularity

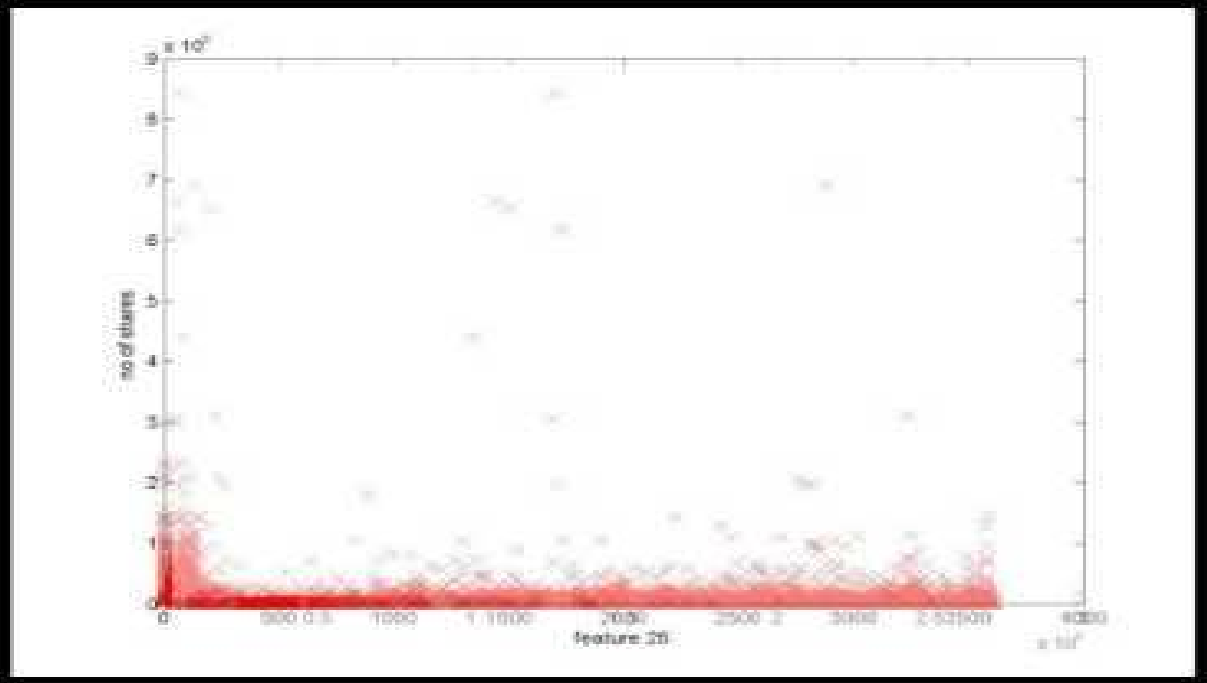
- Set of features about articles published by Mashable in two years.
- Number of instances- 39k
- 58 predictive fields, 1 goal field(no of shares)
- No of words, no of images, etc are few among the predictive features.

- Using ML, “no of shares” would be predicted using different techniques.
- Datset gathered from https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity

Techniques Implemented

- Linear Decision Boundary
  - Linear Regression using Gradient Descent Method
  - Linear Regression using Normal Method.
- Polynomial Decision Boundary
  - Gradient Descent Method.
  - Normal Method.
- Logistic Regression

Visualaisation of Data



Gradient Descent Method

- Hypothesis Function
- Cost Function

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

- Feature Scaling
- Gradient Descent Convergence

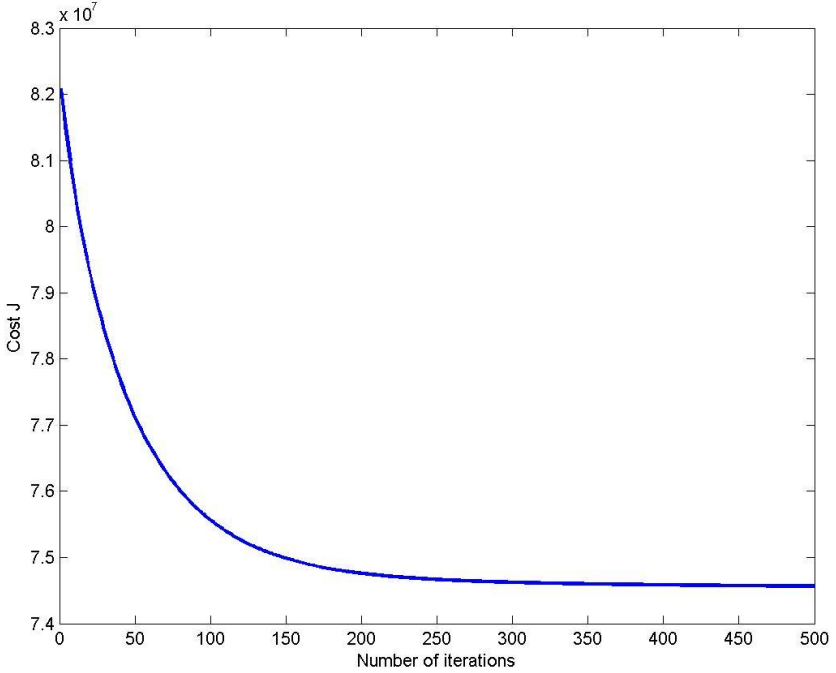
repeat until convergence: {

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)} \qquad \text{for j := 0..n}$$

}

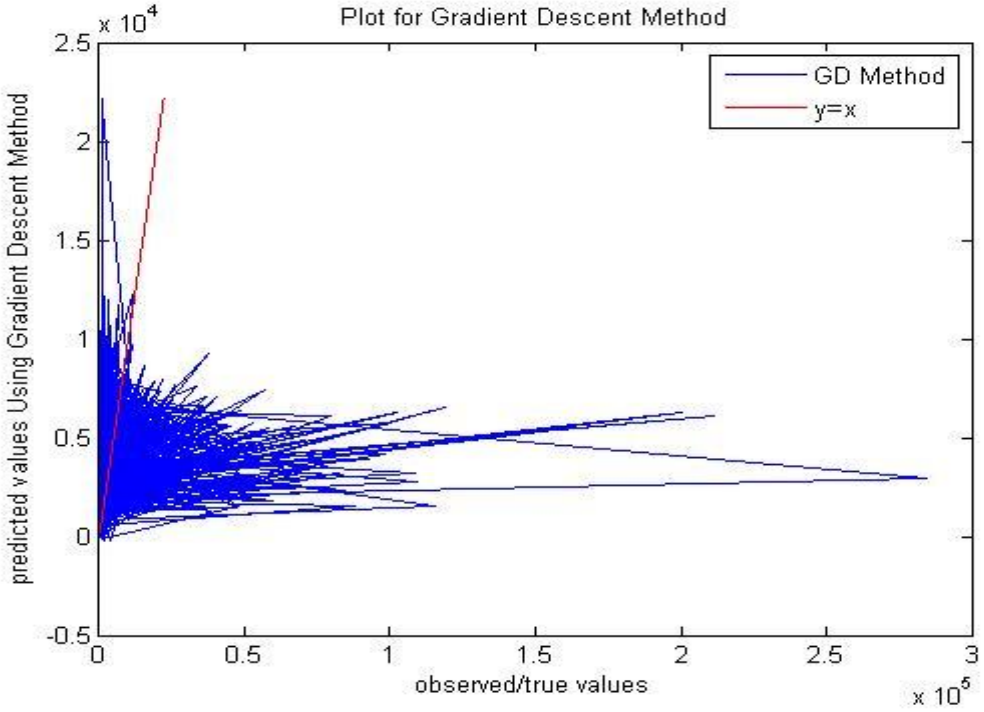
Cost Function for gradient Descent Method

- Cost Function decreases with iterations.
- Learning rate, alpha= 0.03



Prediction(GD Method)

- Training Accuracy- close to zero
- Accuracy is found to be approx 6% when 10% of shift in predicted values are acceptable.
- When 20% of shift is acceptable, accuracy is approx 12%



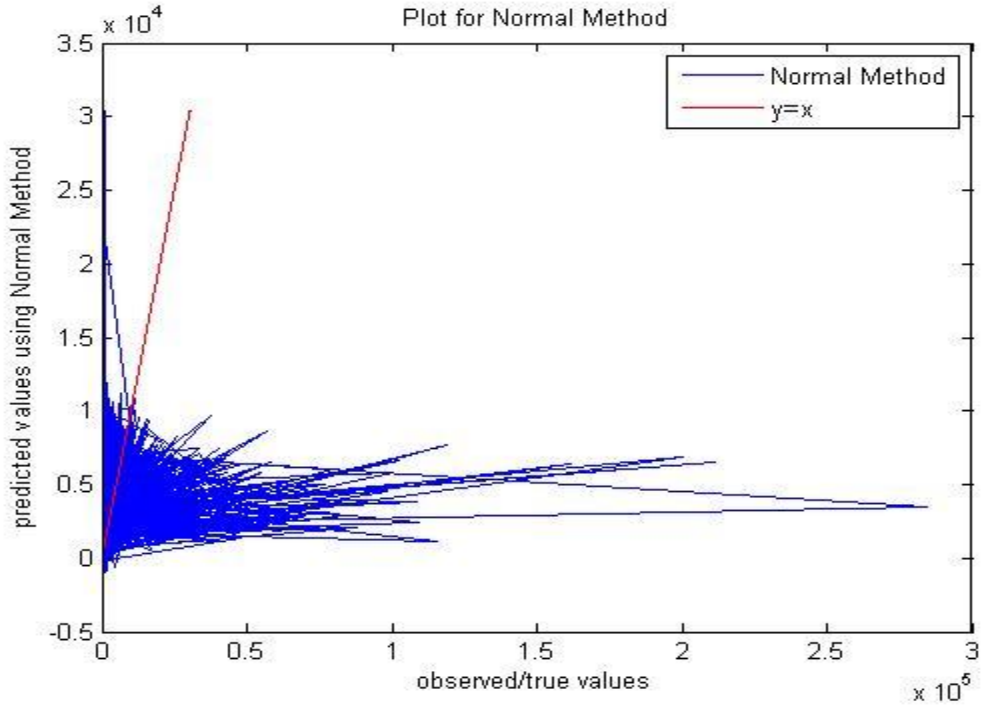
Normal Method

- Method of finding optimum theta without iterations.
- No need of feature scaling
-

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

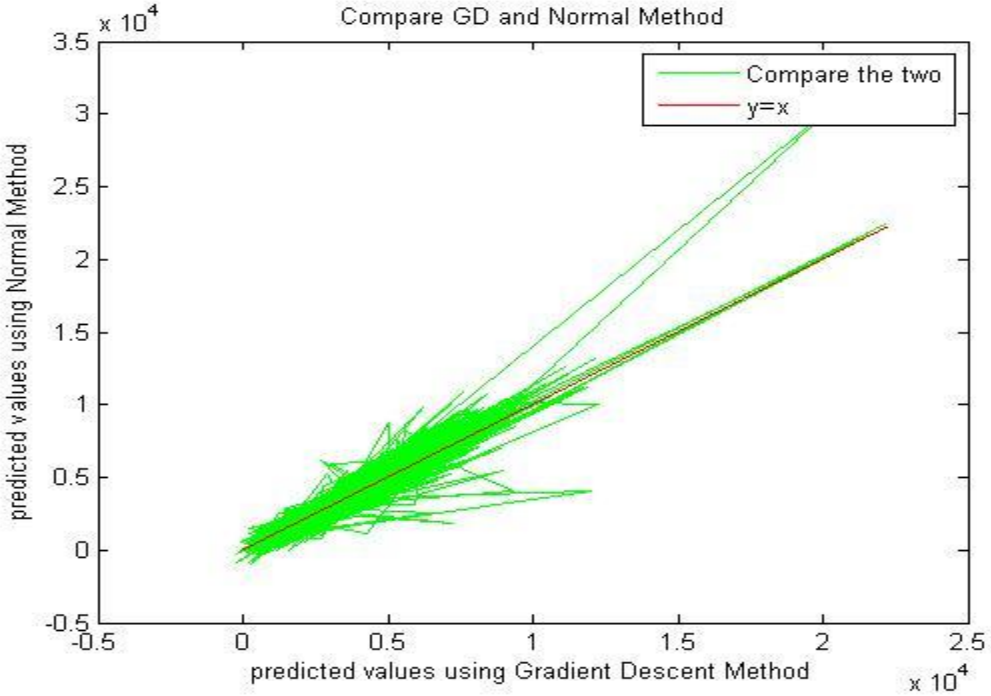
Prediction(Normal Method)

- Training Accuracy close to zero.
- Accuracy is found to be approx 7% when 10% of shift in predicted values are acceptable.
- When 20% of shift is acceptable, accuracy is approx 13%



Compare GD and Normal Method Results

- Similarity in the results by 2 methods can be seen here.
- Several “theta” or weight values were almost same in the two cases.

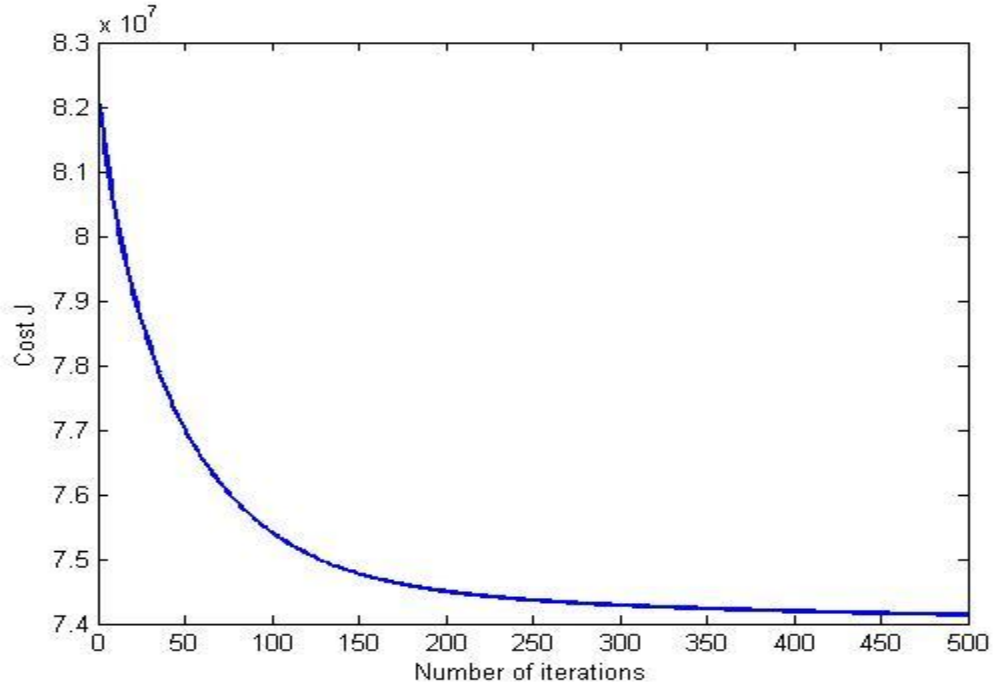


Polynomial Regression

- Hypothesis function need not be linear, if it doesn’t fit the data well.
- We can change the curve of our hypothesis by making it quadratic, cubic or any other form.

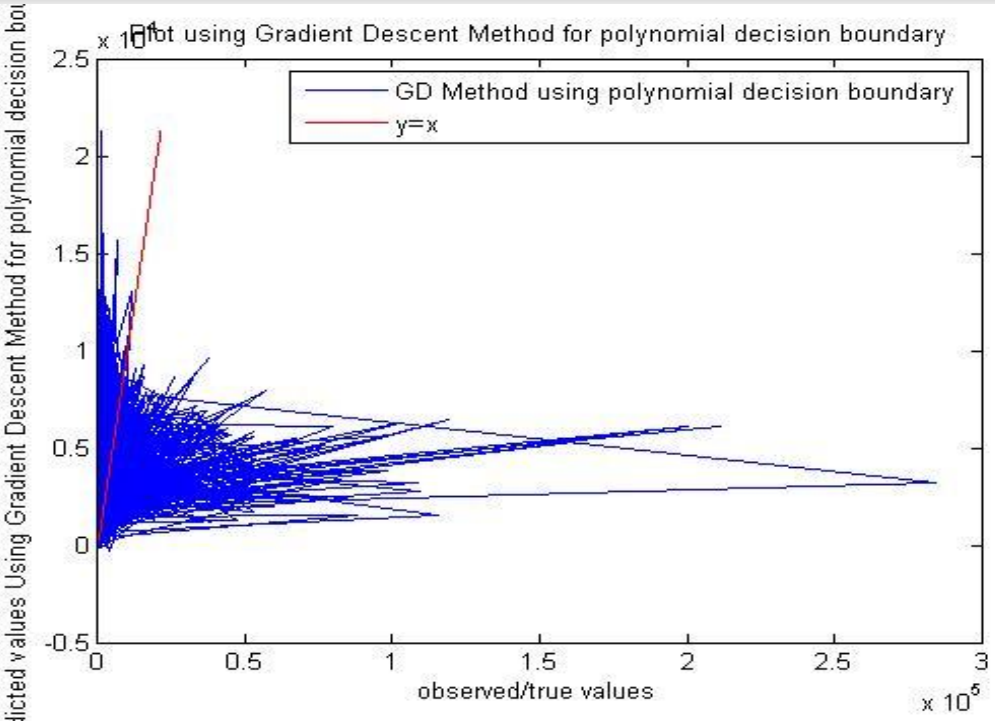
Cost Function(GD) for Polynomial Regression

- Cost Function Decreasing
- Learning Rate, alpha= 0.01



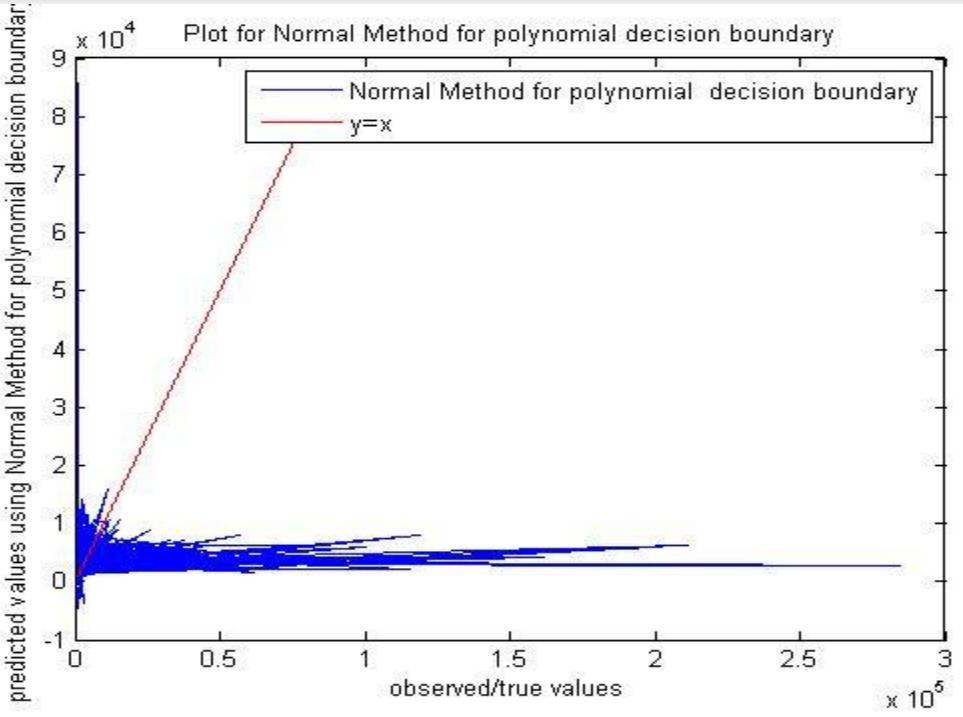
Prediction(GD) for Polynomial Regression

- Training Accuracy close to zero.
- Accuracy is found to be approx 6% when 10% of shift in predicted values are acceptable.
- When 20% of shift is acceptable, accuracy is approx 12%

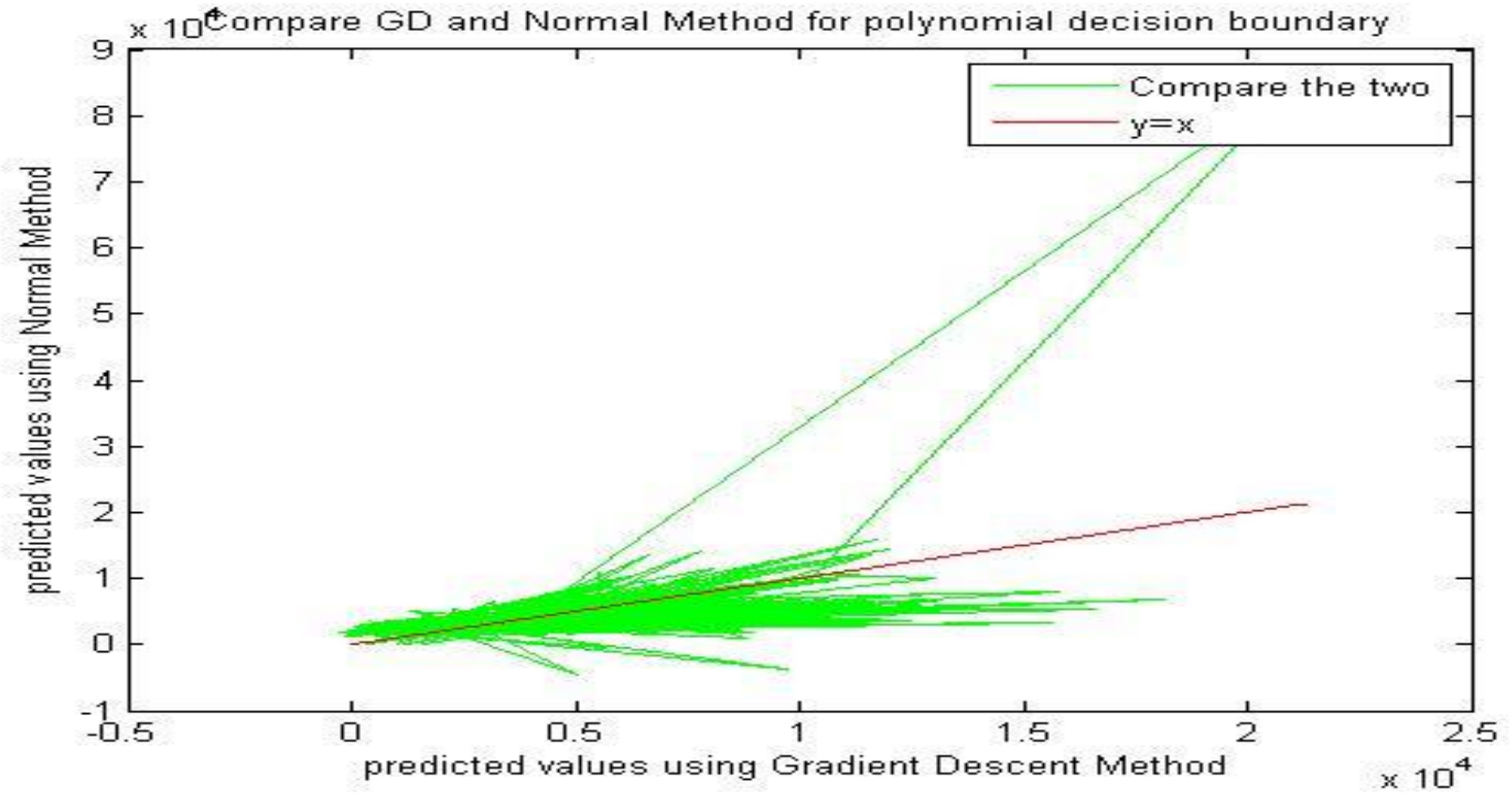


Prediction(Normal) for Polynomial Regression

- Training Accuracy close to zero.
- Accuracy is found to be approx 5% when 10% of shift in predicted values are acceptable.
- When 20% of shift is acceptable, accuracy is approx 9%



Prediction for GD and Normal Method for Polynomial Regression



Conclusion

- These algorithms are not working well enough to predict the goal output “practically” correct.
- Need to search for other algorithms
- Possibly, many more suitable instances required to learn properly.
- Manipulation of features are necessary.

Fututre Works

- Finding the most effective algorithm to predict the number of shares and studying the behaviour of these algorithms.
- Applying several techniques such as Neural Network, etc
- Different View :- Seeing the “number of shares” as discrete values, apply the classification algos and compare it with regression results.
- Creating a predictor of graph behaviour of different features and utilising it to form hypothesis function.

References

- https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity
- http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_158.pdf
- http://www.nedsi.org/proc/2007/proc/p061011026.pdf
- Huge thanks to Andrew Ng for his course materials on Coursera.

Thank you