University of Information Technology and Sciences (UITS)
Department of Information Technology

Lab Report No.: 1

IT-452 : Machine Learning

Linear, Polynomial Multiple

Submitted To:

Sumaiya Akhtar Mitu
Lecturer,
Department of IT ,
UITS
Email:sumaiya.akhtar@uits.edu.bd

Submitted By:

Name:Nazmul Zaman
Student ID:2014755055
Department of IT, UITS

Contents

1 Abstract **2**

2 Linear Regression **2**

3 Code **2**

4 Polynomial Regression **4**

5 Advantage of Regression **6**

6 Disadvantages of Regression **6**

7 Conclusion **6**

8 References **6**

1 Abstract

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.Regression analysis includes several variations, such as linear, multiple linear, and nonlinear. The most common models are simple linear and multiple linear. Nonlinear regression analysis is commonly used for more complicated data sets in which the dependent and independent variables show a nonlinear relationship.

2 Linear Regression

Linear regression is a statistical regression method which is used for predictive analysis. It is one of the very simple and easy algorithms which works on regression and shows the relationship between the continuous variables. It is used for solving the regression problem in machine learning. Linear regression shows the linear relationship between the independent variable (X-axis) and the dependent variable (Y-axis), hence called linear regression.

3 Code

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset = pd.read_csv('https://gist.githubusercontent.com/RobotOptimist/
    cc82e87e7d2104e58711b7c846a9e220/raw/69
    ca02a46df0be8c210ebbded3106c5d9956a4bc/Position_Salaries.csv')
X = dataset.iloc[:, 1:-1].values
y = dataset.iloc[:, -1].values

from sklearn.linear_model import LinearRegression
lin_reg = LinearRegression()
lin_reg.fit(X, y)


plt.scatter(X, y, color = 'red')
plt.plot(X, lin_reg.predict(X), color = 'blue')
plt.title('Truth or Bluff (Linear Regression)')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```

Figure 1:

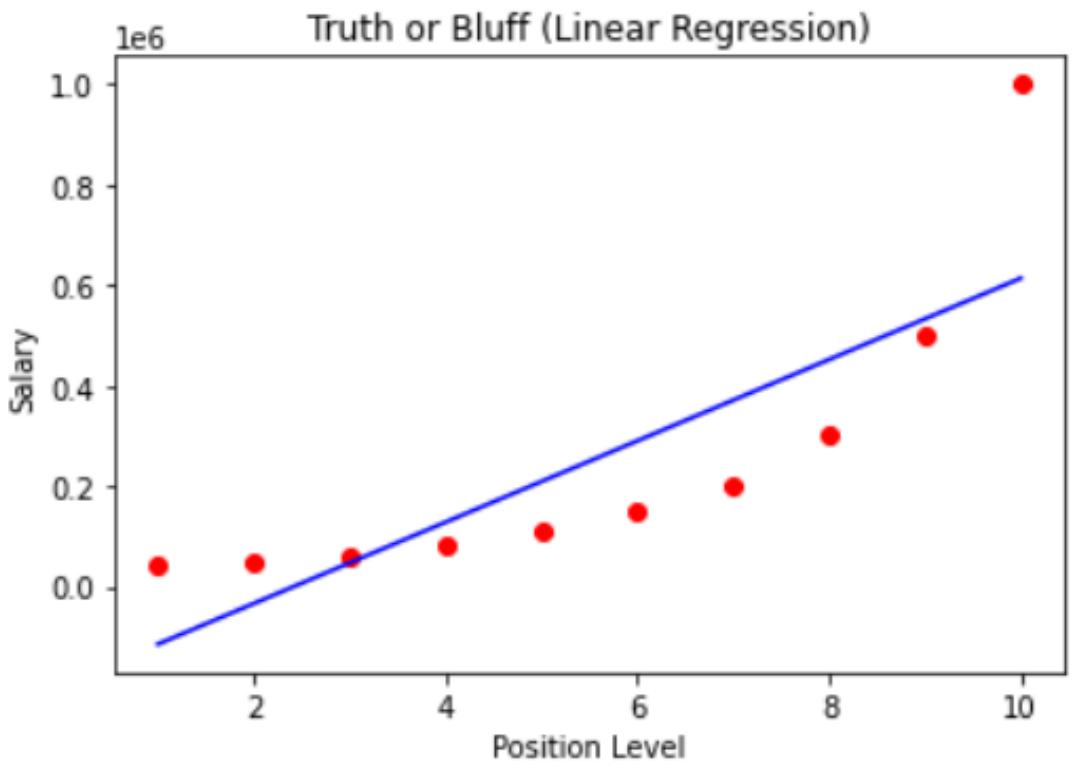

Figure 2: Linear Regression

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
dataset = pd.read_csv('https://gist.githubusercontent.com/RobotOptimist/
    cc82e87e7d2104e58711b7c846a9e220/raw/69
    ca02a46df0be8c210ebbded3106c5d9956a4bc/Position_Salaries.csv')
X = dataset.iloc[:, 1:-1].values
y = dataset.iloc[:, -1].values
```

Figure 3:

4 Polynomial Regression

Polynomial Regression is a type of regression which models the non-linear data set using a linear model. It is similar to multiple linear regression, but it fits a non-linear curve between the value of x and corresponding conditional values of y. Suppose there is a dataset which consists of datapoints which are present in a non-linear fashion, so for such case, linear regression will not best fit to those datapoints. To cover such datapoints, we need Polynomial regression. In Polynomial regression, the original features are transformed into polynomial features of given degree and then modeled using a linear model. Which means the datapoints are best fitted using a polynomial line.

```
from sklearn.preprocessing import PolynomialFeatures
poly_reg = PolynomialFeatures(degree = 10)
X_poly = poly_reg.fit_transform(X)
lin_reg_2 = LinearRegression()
lin_reg_2.fit(X_poly, y)


plt.scatter(X, y, color = 'red')
plt.plot(X, lin_reg_2.predict(poly_reg.fit_transform(X)), color = 'blue')
plt.title('Truth or Bluff (Polynomial Regression)')
plt.xlabel('Position level')
plt.ylabel('Salary')
plt.show()
```

Figure 4:

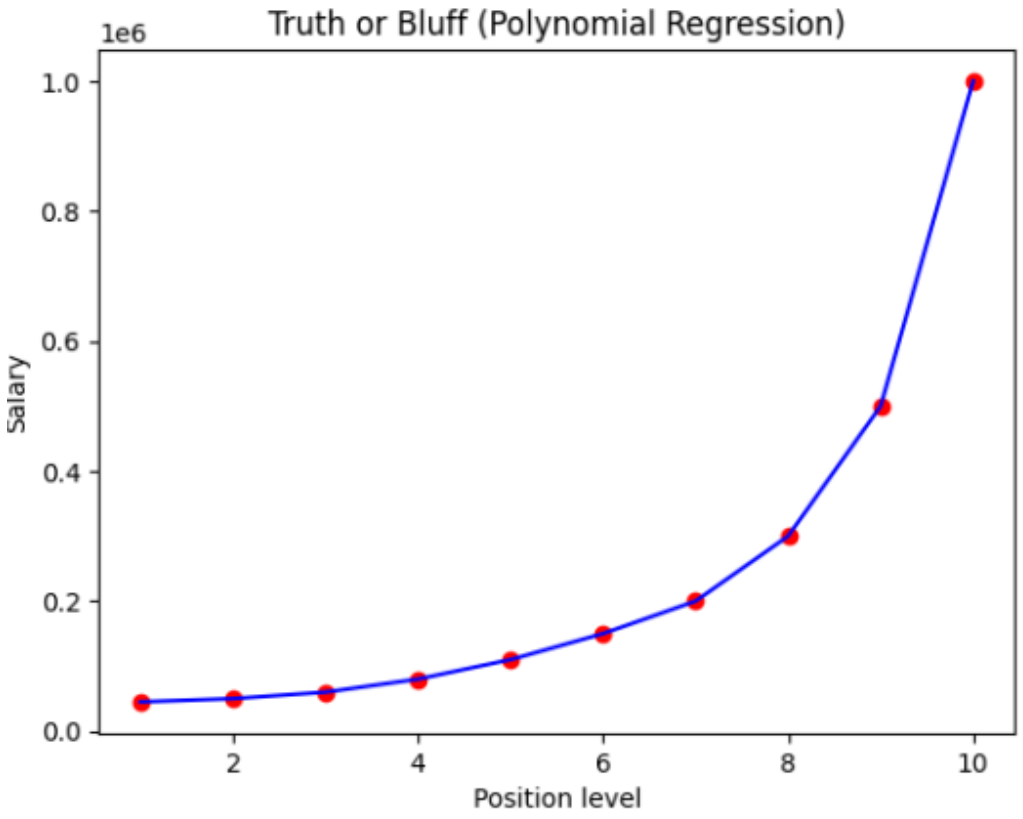

Figure 5: polynomial Regression

5 Advantage of Regression

1. Regression models are easy to understand as they are built upon basic statistical principles, such as correlation and least-square error.
2. The output of regression models is an algebraic equation that is easy to understand and use to predict.
3. The strength (or the goodness of fit) of the regression model is measured in terms of the correlation coefficients, and other related statistical parameters that are well understood.
4. The predictive power of regression models matches with other predictive models and sometimes performs better than the competitive models.
5. Regression models can include all the variables that one wants to include in the model.

6 Disadvantages of Regression

1. Regression models cannot work properly if the input data has errors (that is poor quality data). If the data preprocessing is not performed well to remove missing values or redundant data or outliers or imbalanced data distribution, the validity of the regression model suffers.
2. Regression models are susceptible to collinear problems (that is there exists a strong linear correlation between the independent variables). If the independent variables are strongly correlated, then they will eat into each other's predictive power and the regression coefficients will lose their ruggedness.
3. As the number of variables increases the reliability of the regression models decreases. The regression models work better if you have a small number of variables.
4. Regression models do not automatically take care of non linearity. The user needs to imagine the kind of additional terms that might be needed to be added to the regression model to improve its fit.

7 Conclusion

Machine learning analytics uses machine learning regression models mostly to forecast trends and forecast results. To comprehend the relationship between several independent variables and a result, regression models will be trained.

8 References

2. https://www.javatpoint.com/linear-regression-in-machine-learning
3. https://www.javatpoint.com/machine-learning-polynomial-regression