



Ai and Machine Learning Option 2

Project Report (title): Linear and Polynomial Regression

Group member: Md Amdadul Hoque (2004103)

Md Forkan Hossain (1902394)

Algorithm/Model: bellow,

- Linear Regression
- Polynomial Regression

Description

- The dataset has information about persons age, income, and other information regarding buying a house.
- The goal is to predict house pricing.

Data preparation

- The dataset has 7 columns and 5000 entries for each column.
- The dataset has been divided in 30% test and 70% train.

- | | Avg. Area li | Avg. Area l | Avg. Area l | Avg. Area l | Area Popul | Price | Address |
|----|--------------|-------------|-------------|-------------|------------|----------|-----------|
| 2 | 79545.46 | 5.682861 | 7.009188 | 4.09 | 23086.8 | 1059034 | 208 |
| 3 | 79248.64 | 6.0029 | 6.730821 | 3.09 | 40173.07 | 1505891 | 188 |
| 4 | 61287.07 | 5.86589 | 8.512727 | 5.13 | 36882.16 | 1058988 | 9127 |
| 5 | 63345.24 | 7.188236 | 5.586729 | 3.26 | 34310.24 | 1260617 | USS |
| 6 | 59982.2 | 5.040555 | 7.839388 | 4.23 | 26354.11 | 630943.5 | USNS |
| 7 | 80175.75 | 4.988408 | 6.104512 | 4.04 | 26748.43 | 1068138 | 06039 |
| 8 | 64698.46 | 6.025336 | 8.14776 | 3.41 | 60828.25 | 1502056 | 4759 |
| 9 | 78394.34 | 6.98978 | 6.620478 | 2.42 | 36516.36 | 1573937 | 972 Joyce |
| 10 | 59927.66 | 5.362126 | 6.393121 | 2.3 | 29387.4 | 798869.5 | USS |
- There two model have been used. Linear and Polynomial regression

Relevant metrics

- Here several metrics have been used.
 - I. MAE and RMSE: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are metrics used to evaluate a Regression Model.
 - II. MSE: The MSE is a measure of the quality of an estimator.
 - III. R2 is a measure of the goodness of fit of a model.

- ```
13 def print_evaluate(true, predicted):
14 mae = metrics.mean_absolute_error(true, predicted)
15 mse = metrics.mean_squared_error(true, predicted)
16 rmse = np.sqrt(metrics.mean_squared_error(true, predicted))
17 r2_square = metrics.r2_score(true, predicted)
18 print('MAE:', mae)
19 print('MSE:', mse)
20 print('RMSE:', rmse)
21 print('R2 Square', r2_square)
22 print('_____')
```

## Conclusion

- I run the model 3 times and I got satisfactory results to me and my team member.
- Here is the accuracy for each model used here.

|   | A | B          | C        | D        | E        | F         |
|---|---|------------|----------|----------|----------|-----------|
| 1 |   | Model      | MAE      | MSE      | RMSE     | R2 Square |
| 2 | 0 | polynomial | 81174.52 | 1.01E+10 | 100409.1 | 0.914567  |
| 3 | 1 | linear     | 81135.57 | 1.01E+10 | 100341.5 | 0.914682  |

- And here is result of 3 runtimes

|   | A | B            | C           | D           | E           | F          | G       |
|---|---|--------------|-------------|-------------|-------------|------------|---------|
| 1 |   | Avg. Area li | Avg. Area l | Avg. Area l | Avg. Area l | Area Popul | Price   |
| 2 | 0 | 79545.46     | 5.682861    | 7.009188    | 4.09        | 23086.8    | 1220433 |
| 3 | 1 | 79545        | 5           | 7           | 4           | 23086      | 1105010 |
| 4 | 2 | 79545        | 5           | 7           | 4           | 23086      | 1109032 |

- Finally, we are happy with our mode. But we could add more data to the dataset to get better result.