

IE 345 - K “Introduction to Deep Learning: Fundamentals Concepts”

Prof. Yuzo

Random Forest Regression

pg. 75 - 76

```
In [1]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from numpy import array
```

```
In [2]: dataset = pd.read_csv('Position_Salaries.csv')
x = dataset.iloc[:, 1:2].values
y = dataset.iloc[:, 2].values
dataset.head()
```

Out[2]:

	Position	Level	Salary
0	Business Analyst	1	45000
1	Junior Consultant	2	50000
2	Senior Consultant	3	60000
3	Manager	4	80000
4	Country Manager	5	110000

```
In [3]: # Splitting the dataset into the training and test set
from sklearn.model_selection import train_test_split

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)
```

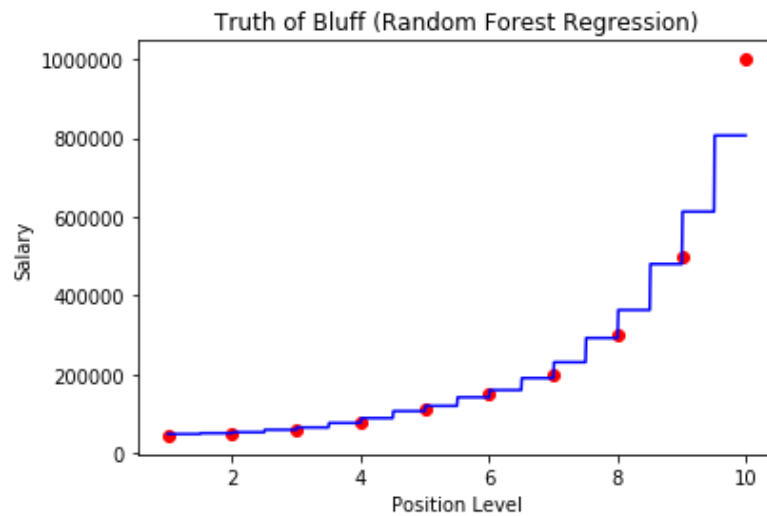
```
In [5]: # Feature Scaling
from sklearn.preprocessing import StandardScaler
sc_x = StandardScaler()
x_train = sc_x.fit_transform(x_train)
x_test = sc_x.transform(x_test)
sc_y = StandardScaler()
y_train = sc_y.fit_transform(y_train.reshape(-1, 1)) #recommendation of python
```

```
In [6]: # Fitting Random Forest Regression to the Dataset
from sklearn.ensemble import RandomForestRegressor

regressor = RandomForestRegressor(n_estimators = 300, random_state = 0)
regressor.fit(x, y)

# Predicting a new result
y_pred = regressor.predict(array(6.5).reshape(-1, 1))
# Recommendation of Python take of https://www.kaggle.com/pratjain/
```

```
In [7]: # Visualising the Random Forest Regression
x_grid = np.arange(min(x), max(x), 0.01)
x_grid = x_grid.reshape((len(x_grid), 1))
plt.scatter(x, y, color='red')
plt.plot(x_grid, regressor.predict(x_grid), color='blue')
plt.title('Truth of Bluff (Random Forest Regression)')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```



Pablo David Minango Negrete

pablodavid218@gmail.com

Lisber Arana Hinostroza

lisberarana@gmail.com