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

Prof. Yuzo

Random Forest Regression

pg. 75 - 76

In [3]:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

In [12]:

```
dataset = pd.read_csv('C:/Users/pablo/Desktop/Pablo David/UNICAMP/Python/IE345-K_DeepLearning/Position_Salaries.csv')
x = dataset.iloc[:, 1:2].values
y = dataset.iloc[:, 2].values
print('X' ,x)
print('Y' ,y)

# 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)
```

```
X [[ 1]
 [ 2]
 [ 3]
 [ 4]
 [ 5]
 [ 6]
 [ 7]
 [ 8]
 [ 9]
 [10]]
Y [  45000   50000   60000   80000  110000  150000  200000  300000  500000
 1000000]
```

In [13]:

```
# 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)) #recomendation of python
```

```
C:\Users\pablo\Anaconda3\lib\site-packages\sklearn\utils\validation.py:47
5: DataConversionWarning: Data with input dtype int64 was converted to flo
at64 by StandardScaler.
  warnings.warn(msg, DataConversionWarning)
```

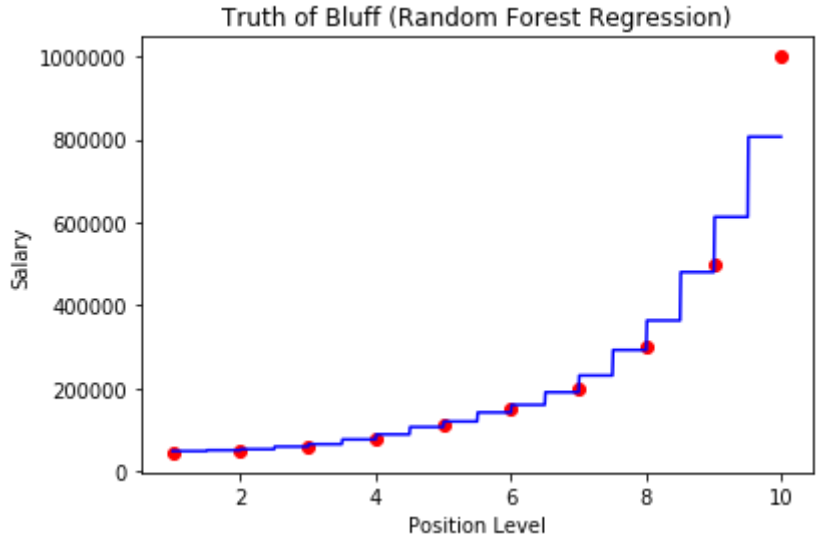
In [16]:

```
# 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(6.5)
```

In [17]:

```
# 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