

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

Prof. Yuzo

Decision Tree

pg. 72 - 74

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

```
In [2]: # Importing the dataset
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)
```

Note: Library `sklearn.cross_validation` change for `sklearn.model_selection`, for more information visit:

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html (https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html)

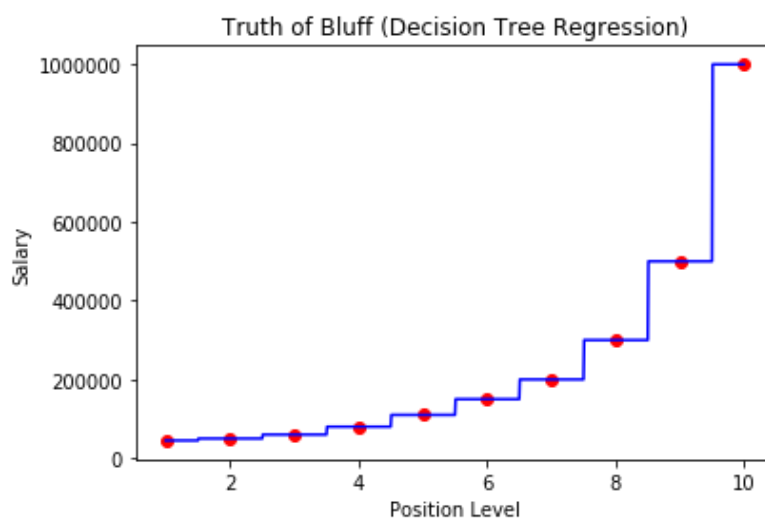
```
In [4]: # Fitting Decision Tree Regression to the dataset
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state = 0)
regressor.fit(x, y)
```

```
Out[4]: DecisionTreeRegressor(criterion='mse', max_depth=None, max_features=None,
                               max_leaf_nodes=None, min_impurity_decrease=0.0,
                               min_impurity_split=None, min_samples_leaf=1,
                               min_samples_split=2, min_weight_fraction_leaf=0.0,
                               presort=False, random_state=0, splitter='best')
```

```
In [5]: # Predicting a new result
y_pred = regressor.predict(array(6.5).reshape(-1, 1))
# Recommendation of Python take of https://www.kaggle.com/pratjain/decision-tree-regression
```

```
In [6]: # Visualising the Decision Tree 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 (Decision Tree Regression)')
plt.xlabel('Position Level')
plt.ylabel('Salary')
plt.show()
```



Pablo David Minango Negrete

pablodavid218@gmail.com

Lisber Arana Hinostroza

lisberarana@gmail.com