Decision Tree Regression

Rayan Crasta (stea1th9)

Problem Statement

Lets imagine we are an HR and we ant to hire , we found a great fit for the job . But the question comes what is your salary expectiation he demands $160,000 per year in prev company We are gonna build a plynomial regression to predict his previous salary to know wether its the truth of the bluff:

Data Set:

- **Positions**
- **Level**
- **Salary**

We need to check which position this person had according to salary .

But chance is he has been Regional Manager for Dunder Mifflin for quite a while. Hence his salary shouldnt be 150,000 but it should be between 150,000 to 160,000 i.e position level 6 and 7

Importing the libraries

In [0]:

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

Importing the dataset

In [0]:

```
dataset = pd.read_csv('Position_Salaries.csv')
X=dataset.iloc[:,1:-1].values
#select only level column cuz its like the  encoded version of position
y = dataset.iloc[:,-1].values
```

Training the Decision Tree Regression model on the whole dataset

In [3]:

```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor(random_state=0)
regressor.fit(X,y)
```

Out[3]:

```
DecisionTreeRegressor(ccp_alpha=0.0, 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='deprecated',
                      random_state=0, splitter='best')
```

Predicting a new result

In [5]:

```
regressor.predict([[6.5]])
#prediction is not bas as the prediction is lower than the demanded salary
```

Out[5]:

```
array([150000.])
```

Visualising the Decision Tree Regression results (higher resolution)

In [6]:

```
#as we have small Data set ,we will increase x vals plots
X_grid= np.arange(min(X),max(X),0.1)#create array at differnece of 0.1 cuz small dataset
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 or Bluff(Decision Tree)')
plt.xlabel('Position Level')
plt.ylabel("Salary")
plt.show()

#Regression tree is not very suitable to 2D dataset(1 dependent and 1 independent)
#The result isnt beautiful though cuz the model to each salary from positon level 0.5 and
-0.5 as predicted the salary
#to be the same as the moddle hence the flat lines
#as data is split at succesive nodes
```