Random Forest 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 matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

In [0]:

```
dataset = pd.read_csv('Position_Salaries.csv')
X = dataset.iloc[:, 1:-1].values
y = dataset.iloc[:, -1].values
```

Training the Random Forest Regression model on the whole dataset

In [0]:

```
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators = 10, random_state = 0) #10 trees (each tree is an estimator)
regressor.fit(X, y)
```

Out[0]:

```
RandomForestRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse',
                      max_depth=None, max_features='auto', max_leaf_nodes=None,
                      max_samples=None, min_impurity_decrease=0.0,
                      min_impurity_split=None, min_samples_leaf=1,
                      min_samples_split=2, min_weight_fraction_leaf=0.0,
                      n_estimators=10, n_jobs=None, oob_score=False,
                      random_state=0, verbose=0, warm_start=False)
```

Predicting a new result

In [0]:

```
regressor.predict([[6.5]])
```

Out[0]:

```
array([167000.])
```

Visualising the Random Forest Regression results (higher resolution)

In [0]:

```
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 or Bluff (Random Forest Regression)')
plt.xlabel('Position level')
plt.ylabel('Salary')
plt.show()

#Output isnt beautiful because the Rnadom Forest Regression is used to high dimensional d
ataset
```