Random_forest_and_extra_trees

March 10, 2019

```
In [1]: import numpy as np
        from sklearn.model_selection import cross_val_score
        from sklearn.ensemble import RandomForestClassifier, ExtraTreesClassifier, RandomForest
```

```
In [2]: # Classification
        import pickle
        import numpy as np
        covertype_dataset = pickle.load(open('covertype_dataset.pickle','rb'))
        covertype_X = covertype_dataset.data[:80000,:]
        covertype_Y = covertype_dataset.target[:80000] -1
        covertypes = ['Spruce/Fir','Lodgepole Pine', 'Ponderosa Pine', 'Cottonwod/Wollow', 'Asp
```

```
In [3]: %%time
        hypothesis = RandomForestClassifier(n_estimators=600, random_state=101)
        scores = cross_val_score(hypothesis, covertype_X, covertype_Y, cv=3, scoring='accuracy
        print("RandomForestClassifier -> accuracy of cross-validation:\nmean = %f\nstandard de
```

```
RandomForestClassifier -> accuracy of cross-validation:
mean = 0.889075
standard deviation = 0.000290
Wall time: 1min 13s
```

```
In [4]: %%time
        hypothesis = ExtraTreesClassifier(n_estimators=600, random_state=101)
        scores = cross_val_score(hypothesis, covertype_X, covertype_Y, cv=3, scoring='accuracy
        print("ExtraTreesClassifier -> accuracy of cross-validation:\nmean = %f\nstandard devia
```

```
ExtraTreesClassifier -> accuracy of cross-validation:
mean = 0.895587
standard deviation = 0.000729
Wall time: 1min 34s
```

```
In [5]: # Regression
        import pickle
        X_train, y_train = pickle.load(open('cadata.pickle','rb'))
        from sklearn.preprocessing import scale
```

```
        first_rows = 6000
        X_train = scale(X_train[:first_rows,:].toarray())
        y_train = y_train[:first_rows]/10**4.0 # Resoults will be in 1000s of dolars

In [6]: hypotesis = RandomForestRegressor(n_estimators=600, random_state=101)
        scores = cross_val_score(hypotesis, X_train, y_train, cv=3, scoring='neg_mean_absolute_
        print("RandomForestRegressor -> accuracy of cross-validation:\nmean = %f\nstandard devi

RandomForestRegressor -> accuracy of cross-validation:
mean = -8.373486
standard deviation = 0.612176


In [7]: hypotesis = ExtraTreesRegressor(n_estimators=600, random_state=101)
        scores = cross_val_score(hypotesis, X_train, y_train, cv=3, scoring='neg_mean_absolute_
        print("RandomForestRegressor -> accuracy of cross-validation:\nmean = %f\nstandard devi

RandomForestRegressor -> accuracy of cross-validation:
mean = -8.129034
standard deviation = 0.700811


In [ ]:
```