

3_random_forest_exercise

May 31, 2019

```
In [ ]: import numpy as np

In [ ]: X = np.load("./tatanic_X_train.npy")
        y = np.load("./tatanic_y_train.npy")

In [ ]: from sklearn.model_selection import train_test_split
        X_train, X_test, y_train, y_test = train_test_split(X, y, \
                                                            test_size=0.3, random_state=101)

In [ ]: X_train[0]

In [ ]: y_train[:10]

In [ ]: from sklearn.ensemble import RandomForestClassifier
        eclf = RandomForestClassifier(n_estimators=100, max_features=2, \
                                      n_jobs=7, oob_score=True)

In [ ]: from sklearn.model_selection import cross_val_score
        cross_val_score(eclf, X_train, y_train, cv=5).mean()

In [ ]: params ={
        "n_estimators" : [10, 20, 30, 50, 100],
        "max_features" : [1,2,3,4,5,6,7, 10, 15, 20, 25, len(X[0])]
        }

In [ ]: # cross_val_score - cross validation
        # GridSearchCV - cross validation, parm grid search, model return

        # gridsearchcv operation

In [ ]: grid.best_score_

In [ ]: grid.best_params_

In [ ]: grid.best_estimator_.oob_score_

In [ ]: grid.best_estimator_

In [ ]:
```

```
In [ ]: grid.score(X_test, y_test)

In [ ]: # confusion matrix

In [ ]: # classification report

In [ ]: btree = RandomForestClassifier(max_features = 20, n_estimators = 20)

In [ ]: btree.fit(X_train, y_train)
        btree.score(X_test, y_test)

In [ ]: btree.feature_importances_
```