Classification reports

January 29, 2016

In [1]:
```
from predict import prepare_sets

data, sub_data, survived = prepare_sets()
test_data = sub_data[data.type == 'test']
train_data = sub_data[data.type == 'train']
```

In [*]:
```
from sklearn.cross_validation import train_test_split
from sklearn.grid_search import GridSearchCV
from sklearn.metrics import classification_report
from sklearn import ensemble

X_train, X_test, y_train, y_test = train_test_split(
    train_data, survived, test_size=0.2, random_state=0)

params = [{
    'n_estimators': [100, 300, 1000],
    'min_density': [0.1, 0.3, 0.5, 0.7, 0.9],
    'max_depth': [None, 1, 2, 5, 10, 20, 50, 100],
    'min_samples_leaf': [1, 2, 5, 10],
    'min_samples_split': [2, 4, 8, 10, 20],
    'max_features': ['auto', 'sqrt', 'log2', None]
}]

clf = ensemble.RandomForestClassifier(n_estimators=10, random_state=1)
clf = GridSearchCV(clf, params, cv=3, n_jobs=3)
clf.fit(X_train, y_train)

print clf.best_estimator_

y_true, y_pred = y_test, clf.predict(X_test)
print classification_report(y_true, y_pred)
```