

CSE4020 Machine Learning
Lab Exercise 7 - Ensemble Learning

Harshit Kedia
15BCE0329
Slot - L19 + L20

Ensemble Learning

Implement AdaBoost algorithm to boost the performance of decision trees on binary classification problems

Dataset: Diabetes data set

<https://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-diabetes.data>

Code:

```
# AdaBoost Classification
import pandas
from sklearn import model_selection
from sklearn.ensemble import AdaBoostClassifier
url =
"https://archive.ics.uci.edu/ml/machine-learning-databases/pima-indians-diabetes/pima-indians-
diabetes.data"
names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']
dataframe = pandas.read_csv(url, names=names)
array = dataframe.values
X = array[:,0:8]
Y = array[:,8]
seed = 7
num_trees = 30
kfold = model_selection.KFold(n_splits=10, random_state=seed)
model = AdaBoostClassifier(n_estimators=num_trees, random_state=seed)
results = model_selection.cross_val_score(model, X, Y, cv=kfold)
print(results.mean())
```

Running the example provides a mean estimate of classification accuracy.
0.76045796309

