Bagging_with_week_clarifiers

March 10, 2019

In [1]:
```
# Bootstrap aggregating, also called bagging, is a machine learning ensemble meta-algo
# designed to improve the stability and accuracy of machine learning algorithms used i
# classification and regression. It also reduces variance and helps to avoid overfitti
# Although it is usually applied to decision tree methods, it can be used with any typ
# Bagging is a special case of the model averaging approach.
# https://en.wikipedia.org/wiki/Bootstrap_aggregating

import pickle
import numpy as np
covertype_dataset = pickle.load(open('covertype_dataset.pickle','rb'))
covertype_X = covertype_dataset.data[:50000,:]
covertype_Y = covertype_dataset.target[:50000] -1
covertypes = ['Spruce/Fir','Lodgepole Pine', 'Ponderosa Pine', 'Cottonwod/Wollow', 'Asp
```

In [2]:
```
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import BaggingClassifier
from sklearn.neighbors import KNeighborsClassifier
hypothesis = BaggingClassifier(KNeighborsClassifier(n_neighbors=1),
max_samples=0.7, max_features=0.7, n_estimators=100)
scores = cross_val_score(hypothesis, covertype_X, covertype_Y, cv=3, scoring='accuracy
print("BaggingClassifier -> accuracy of cross-validation:\nmean = %f\nstandard deviati
```

```
BaggingClassifier -> accuracy of cross-validation:
mean = 0.869480
standard deviation = 0.002478
```

In [ ]: