8. Random Forest

March 28, 2022

```
[1]: #import collections
     #l = ['a', 'a', 'a', 'a', 'b', 'c', 'c']
     #c = collections.Counter(l)
     #print(c.most_common())           # [('a', 4), ('c', 2), ('b', 1)]
     #print(c.most_common(1))          # [('a', 4)]
     #print(c.most_common(1)[0])       # ('a', 4)
     #print(c.most_common(1)[0][0])    # a
```

```
[2]: import numpy as np
     from collections import Counter
     from sklearn import datasets
     from sklearn.model_selection import train_test_split
     from sklearn.tree import DecisionTreeClassifier
```

```
[3]: data = datasets.load_breast_cancer()
     X = data.data
     y = data.target
     print(X.shape)
     X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
     # 455 training, 114 test samples
```

```
(569, 30)
```

```
[4]: def sample(X, y):
         n_samples = X.shape[0]
         # from 455 samples, generate 455 samples with replacement
         idxs = np.random.choice(n_samples, n_samples, replace=True)
         return X[idxs], y[idxs]

     def most_common_label(y):
         counter = Counter(y)
         most_common = counter.most_common(1)[0][0]
         return most_common
```

```
[5]: def RF_classifier(X, y, n_trees):
         trees = []
         for _ in range(n_trees):
             tree = DecisionTreeClassifier()
```

```
        X_samp, y_samp = sample(X, y)
        tree.fit(X_samp, y_samp)    # DecisionTreeClassifier.fit()
        trees.append(tree)

    tree_preds = np.array([tree.predict(X) for tree in trees])
    tree_preds = np.swapaxes(tree_preds, 0, 1)
    y_pred = [most_common_label(tree_pred) for tree_pred in tree_preds]
    return np.array(y_pred)
```

[6]:
```
y_pred = RF_classifier(X_train, y_train,n_trees=3)
accuracy = np.sum(y_train == y_pred)/len(y_train)
print("Training Accuracy: ", accuracy)
```

```
Training Accuracy:  0.9934065934065934
```

[ ]: