Machine Learning Assignment #1

CAU SW 20184286 Donghwa Lee

2023.04.21

First, 'show_accuracy' method was created to display the accuracy of each model.

```
def show_accuracy(y_true, y_pred, model_name):
    return pd.Series({'accuracy':accuracy_score(y_true, y_pred)},
                     name=model_name)
```
✓ 0.0s

1. Decision Tree

Experimental setup_

(Since both MNIST and CIFAR have the same mechanism, I will only use the capture of MNIST here.)

GridSearchCV was used to search proper hyperparameter about min samples split, min samples leaf and max leaf nodes.

Decision Tree

```
params_grid = {'min_samples_split':[2, 5, 10],
               'min_samples_leaf': [1, 2, 4],
               'max_leaf_nodes': [5, 10, None]}
```
✓ 0.0s

Four DecisionTree models which have 3, 6, 9, 12 max depth were created. In addition, cross validation was set k to 5.

depth : 3

```
DT_d3 = DecisionTreeClassifier(max_depth=3)
grid_DT = GridSearchCV(DT_d3,
                       param_grid=params_grid,
                       cv=5)
```
✓ 0.0s

depth : 6

```
DT_d6 = DecisionTreeClassifier(max_depth=6)
grid_DT = GridSearchCV(DT_d6,
                       param_grid=params_grid,
                       cv=5)
```
✓ 0.0s

depth : 9

```
DT_d9 = DecisionTreeClassifier(max_depth=9)

grid_DT = GridSearchCV(DT_d9,
                       param_grid=params_grid,
                       cv=5)
```

✓ 0.0s

depth : 12

```
DT_d12 = DecisionTreeClassifier(max_depth=12)

grid_DT = GridSearchCV(DT_d12,
                       param_grid=params_grid,
                       cv=5)
```

✓ 0.0s

Then, I trained each model with the data. After that, the models started predicting y data when given x data, and I was able to obtain the accuracy using the show_accuracy method that was created at first.

```
grid_DT.fit(MNIST_train_images, MNIST_train_labels)

MNIST_train_labels_pred = grid_DT.predict(MNIST_train_images)
MNIST_test_labels_pred = grid_DT.predict(MNIST_test_images)

MNIST_DT_d3_accuracy = pd.concat([show_accuracy(MNIST_train_labels, MNIST_train_labels_pred, 'MNIST_DT_d3_train'),
                                  show_accuracy(MNIST_test_labels, MNIST_test_labels_pred, 'MNIST_DT_d3_test')],
                                 axis=1)

MNIST_DT_d3_accuracy
```

✓ 5m 10.0s

Result_

We can see that as the depth increases, higher accuracy and longer execution time are required.

	MNIST_DT_d3_train	MNIST_DT_d3_test
accuracy	0.491517	0.4953

	CIFAR_DT_d3_train	CIFAR_DT_d3_test
accuracy	0.23762	0.2394

	MNIST_DT_d6_train	MNIST_DT_d6_test
accuracy	0.73825	0.7415

	CIFAR_DT_d6_train	CIFAR_DT_d6_test
accuracy	0.29588	0.2812

	MNIST_DT_d9_train	MNIST_DT_d9_test
accuracy	0.8661	0.8505

	CIFAR_DT_d9_train	CIFAR_DT_d9_test
accuracy	0.38212	0.3042

	MNIST_DT_d12_train	MNIST_DT_d12_test
accuracy	0.93465	0.8774

	CIFAR_DT_d12_train	CIFAR_DT_d12_test
accuracy	0.51882	0.3036

2. SVM

Experimental setup_

Two SVM models which have were created. (linear and rbf)

Then, I trained each model with the data. After that, the models started predicting y data when given x data, and I was able to obtain the accuracy using the show_accuracy method that was created at first.

linear

```
SVM_linear = svm.SVC(kernel='linear')
```

```
SVM_linear.fit(CIFAR_train_images, CIFAR_train_labels)

CIFAR_train_labels_pred = SVM_linear.predict(CIFAR_train_images)
CIFAR_test_labels_pred = SVM_linear.predict(CIFAR_test_images)

CIFAR_SVM_linear_accuracy = pd.concat([show_accuracy(CIFAR_train_labels, CIFAR_train_labels_pred, 'CIFAR_SVM_linear_train'),
                                       show_accuracy(CIFAR_test_labels, CIFAR_test_labels_pred, 'CIFAR_SVM_linear_test')],
                                       axis=1)

CIFAR_SVM_linear_accuracy
```

rbf

```
SVM_rbf = svm.SVC(kernel='rbf')
```

```
SVM_rbf.fit(CIFAR_train_images, CIFAR_train_labels)

CIFAR_train_labels_pred = SVM_rbf.predict(CIFAR_train_images)
CIFAR_test_labels_pred = SVM_rbf.predict(CIFAR_test_images)

CIFAR_SVM_rbf_accuracy = pd.concat([show_accuracy(CIFAR_train_labels, CIFAR_train_labels_pred, 'CIFAR_SVM_rbf_train'),
                                    show_accuracy(CIFAR_test_labels, CIFAR_test_labels_pred, 'CIFAR_SVM_rbf_test')],
                                    axis=1)

CIFAR_SVM_rbf_accuracy
```

Results_

	MNIST_SVM_linear_train	MNIST_SVM_linear_test
accuracy	0.970733	0.9404

	CIFAR_SVM_linear_train	CIFAR_SVM_linear_test
accuracy	0.57488	0.3755

	MNIST_SVM_rbf_train	MNIST_SVM_rbf_test
accuracy	0.989917	0.9792

	CIFAR_SVM_rbf_train	CIFAR_SVM_rbf_test
accuracy	0.70284	0.5436