정보통신대학원 GITA315: Python Machine Learning

Spring 2022

Homework No. 3

Due 11:59 pm, Sunday, June 12, 2022

Submit the paper **by email** before the deadline. Notice that the delayed submission will entail a cut in evaluation by 30% for each delayed day.

Instruction for the submission of the paper is omitted this time. There is no changes to the one for the previous homeworks 1 and 2.

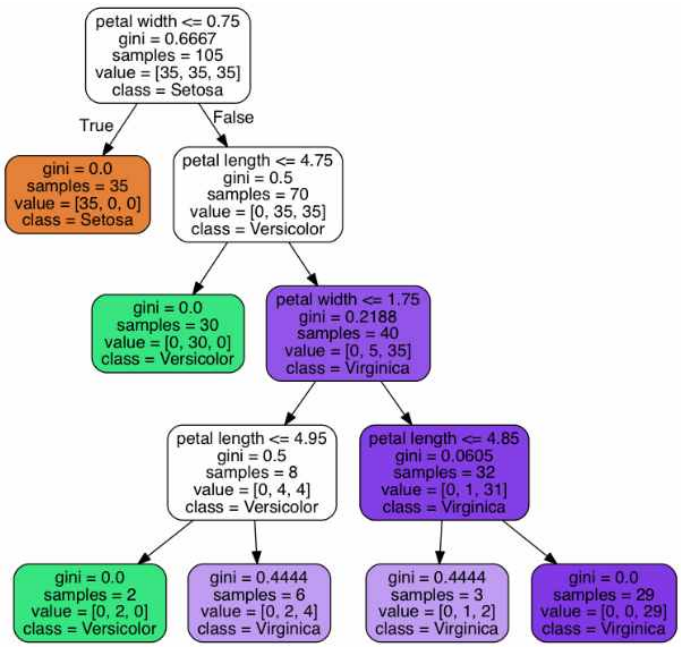
170 points total

Problem 1: 10 points
describe the generalization error.

Problem 2: 20 points total, 10 points each
Considering the visualization (shown below) of the decision tree model discussed in the class,

a) obtain the weighted gini impurity of the chiledren nodes of the parent node whose splitting criterion is petal width <= 1.75.

b) obtain the impurity improvement of the split by that parent node.



Problem 3: 40 points total, 10 points each

a) Describe the reasons why the principal component's directions are highly sensitive to the data scaling in the principal component analysis.

b) Why do we need to get the covariance matrix for PCA?

c) Explain how eigenvector of covariance matrix can be obtained using Python.

d) Describe np.newaxis.

Problem 4: 60 points total

The following code is a part of code for linear discriminant analysis (LDA) algorithm.
Here, S_W is a ($d \times d$) within-class scatter matrix, S_B is a ($d \times d$) between-class scatter matrix.

```
eigen_vals, eigen_vecs = np.linalg.eig(np.linalg.inv(S_W).dot(S_B))

eigen_pairs = [(np.abs(eigen_vals[i]), eigen_vecs[:, i])
               for i in range(len(eigen_vals))]

eigen_pairs = sorted(eigen_pairs, key=lambda k: k[0], reverse=True)
```

Assuming that the above code has been executed,

a) **(10 points)** describe the content in detail and give the shape of `eigen_vals`.

b) **(10 points)** describe the content in detail and give the shape of `eigen_vecs`.

c) **(20 points)** describe the content in detail and give the shape of `eigen_pairs`.

d) **(20 points)** explain the part, `lambda k: k[0]`, in the last line of the above code.

Problem 5: 20 points
Describe the kernel trick.

Problem 6: 20 points
Describe the Grid Search in detail using some sample code of your choice.