

**Final Exam**  
**GITA315: Python Machine Learning**  
**정보통신대학원**  
**June 16, 2022**

- starts at **7:10 pm** and ends at **7:50 pm**
- submit answers **before 7:55 pm**

**Notice:** Answer should be well structured and clear, easy to understand. Answers should be given in complete sentences. Any type of copy and paste should not be included in the answer.

Name: \_\_\_\_\_

Student Number: \_\_\_\_\_

**100 points total**

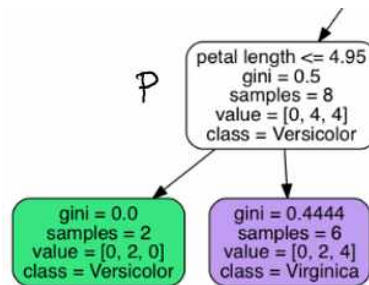
**Problem 1: 15 points total, Python code**

```
w = np.array([1., 0.5, 0.2, 0.2, 0.3, 0.3, 0.5])  
x = np.array([0, 1, 2, 1, 2, 1, 2])
```

- (5 points) What result do you get from `np.bincount(x, weights = w)`.
- (10 points) Explain how the above result is determined.

**Problem 2: 30 points total, 15 points each**

Considering a part of a visualization (shown below) of the decision tree model discussed in the class,



- obtain the weighted gini impurity of the children nodes of the parent node, P, whose splitting criterion is 'petal length <= 4.95'. You should show all the derivations.
- obtain the impurity improvement of the split by that parent node. You should show all the derivations.

**Problem 3: 40 points total, 20 points each**

- Describe the reasons why the principal component's directions are highly sensitive to the data scaling in the principal component analysis.
- Why do we need to get the covariance matrix for PCA?

**Problem 4: 15 points**

Describe the kernel trick.