

# Fundamentals of Data Science

Semester B 20-21

## Tutorial 6

1. a. Instances associated with the left child node of B:  
1(+), 2(+)

Instances associated with the right child node of B  
3(+), 4(-)

Instances associated with the left child node of C  
5(+), 6(+), 7(-), 9(-), 10(-)

Instances associated with the right child node of C  
8(+)

According to the resubstitution estimate, the estimated generalization error rate is

$$\frac{0+1+2+0}{10} = \frac{3}{10} = 0.3$$

- b. Using a penalty term of 0.5 for each leaf node, the estimated generalization error rate becomes

$$\frac{0+1+2+0+0.5(4)}{10} = \frac{5}{10} = 0.5$$

2. a. If the sub-tree associated with node F is pruned and replaced with a leaf node, the estimated generalization error rate is

$$\frac{6+1.5(6)}{24} = \frac{15}{24} = 0.625$$

- b. If the sub-tree associated with nodes D and E are pruned and replaced with leaf nodes, the estimated generalization error rate is

$$\frac{4+1.5(5)}{24} = \frac{11.5}{24} = 0.479$$

- c. If the above operations are performed together, the estimated generalization error rate is

$$\frac{6+1.5(4)}{24} = \frac{12}{24} = 0.5$$