

Tutorial 4

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 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 becomes

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

2. a. If the sub-tree associated with the attribute credit history in the middle branch is replaced with a leaf node, the estimated generalization error is

$$\frac{2+0.6(5)}{14} = \frac{5}{14} = 0.36$$

b. If the sub-tree associated with the attribute credit history in the right branch is replaced with a leaf node, the estimated generalization error is

$$\frac{1+0.6(6)}{14} = \frac{4.6}{14} = 0.33$$

c. If both of the above two operations are performed together, the estimated generalization error is

$$\frac{3+0.6(3)}{14} = \frac{4.8}{14} = 0.34$$