# Tutorial 4

1. Consider the decision tree shown in the following figure:
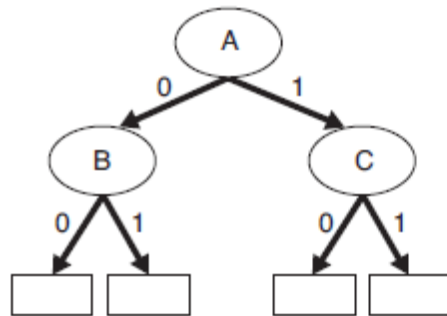


Training:

| Instance | A | B | C | Class |
|----------|---|---|---|-------|
| 1 | 0 | 0 | 0 | + |
| 2 | 0 | 0 | 1 | + |
| 3 | 0 | 1 | 0 | + |
| 4 | 0 | 1 | 1 | − |
| 5 | 1 | 0 | 0 | + |
| 6 | 1 | 0 | 0 | + |
| 7 | 1 | 1 | 0 | − |
| 8 | 1 | 0 | 1 | + |
| 9 | 1 | 1 | 0 | − |
| 10 | 1 | 1 | 0 | − |

   a. Estimate the generalization error of the tree by using the resubstitution estimate.

   b. Estimate the generalization error by using a penalty term of 0.5 for each leaf node.

2. We consider the decision tree constructed for the credit risk estimation problem in the lecture notes. Suppose a penalty term of 0.6 is assigned to each leaf node.

   a. Estimate the generalization error if the sub-tree associated with the attribute **credit history** in the middle branch ($15k to $35k) is replaced with a leaf node.

   b. Estimate the generalization error if, instead, the sub-tree associated with the attribute **credit history** in the right branch (over $35k) is replaced with a leaf node.

   c. Estimate the generalization error if both of the above two operations are performed together.