1. Consider the following training relation:

Salary	Education	Label
10,000	High-school	Reject
40,000	High-school	Reject
40,000	Undergraduate	Accept
15,000	Undergraduate	Reject
75,000	Graduate	Accept
18,000	Graduate	Accept

A decision stump is a decision tree of one node, i.e., only the root. Imagine that we build a decision stump using the above training relation. Compute the 3-fold CV accuracy of the stump if the 3 folds are X = { r_1 , r_3}, Y = { r_2 , r_5} and Z = { r_4, r_6}.

Answer:

(i) Let us build the stump based on X and Y. The training data would be:

Salary	Education	Label
10,000	High-school	Reject
40,000	High-school	Reject
40,000	Undergraduate	Accept
75,000	Graduate	Accept

The stump S_{XY} would consist of the single node Education = High-school. If yes, reject; otherwise, accept.

When applying S_{XY} to predict r_4, r_6, we get 1 correct and 1 incorrect prediction.

(ii) Let us next build the stump based on X and Z. The training data would be:

Salary	Education	Label
10,000	High-school	Reject
40,000	Undergraduate	Accept
15,000	Undergraduate	Reject
18,000	Graduate	Accept

The stump S_{XZ} would consist of the single node salary < 16,500. If yes, reject; otherwise, accept.

When applying S_{XZ} to predict r_2, r_5, we get 1 correct and 1 incorrect prediction.

(iii) Finally, let us build the stump based on Y and Z. The training data would be:

Salary	Education	Label
40,000	High-school	Reject
15,000	Undergraduate	Reject
75,000	Graduate	Accept
18,000	Graduate	Accept

The stump S_{YZ} would consist of the single node Education = Graduate. If yes, aceept; otherwise, reject.

When applying S_{YZ} to predict r_1, r_3, we get 1 correct and 1 incorrect prediction.

Thus, in combining all 3 folds, we get 3 correct and 3 incorrect predictions, for a 3-fold CV accuracy of 0.5.