1. Consider the following training relation:

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

Using entropy as the impurity function, build a decision tree. Show all the calculations involved in finding the winning attributes and the best split points.

Answer:

If Salary $\leq$ 10,000, two subsets $D_1 = \{r_1\}$, D_2 = rest

- $Ent(D_1) = -[1 \log 1 + 0 \log 0] = 0$
- $Ent(D_2) = -[0.75 \log 0.75 + 0.25 \log 0.25] = 0.81$
- $Ent(D) = - [0.6 \log 0.6 + 0.4 \log 0.4] = 0.97$
- Entropy reduction = $Ent(D) – [1/5 *Ent(D_1) + 4/5*Ent(D_2)]= 0.322$

If Salary $\leq$ 15,000, two subsets $D_1 = \{r_{1,} r_3\}$, D_2 = rest

- $Ent(D_1) = -[1 \log 1 + 0 \log 0] = 0$
- $Ent(D_2) = -[1 \log 1 +0 \log 0] = 0$
- $Ent(D) = - [0.6 \log 0.6 + 0.4 \log 0.4] = 0.97$
- Entropy reduction = $Ent(D) – [2/5 *Ent(D_1) + 3/5*Ent(D_2)]= 0.97$.

Because both the entropies of D_1 and D_2 are 0, we split the two classes completely already. No more calculation is needed. The decision tree has a single node, which is salary $\leq$ (15000+18000)/2 = 16,500.