

Assignment 2

By Brandon Young and Ruicheng Wu

1 problem 2

1.1 a.

We claim the provided tree correctly categorize the provided example since every example can be inducted from this decision tree. Like GPA above 3.6 is P and below 3.3 is N, then with publication P, otherwise check University Rank, Only rank 2 will be P, other ranks are N. And recommendation doesn't matter.

1.2 b.

Step I

$$I(\frac{6}{12}, \frac{6}{12}) = 1$$

GPA: $[3.9, 4.0]$ 3(*PPP*) , $(3.2, 3.9)$ 5(*PPPNN*) , $[3.0, 3.2]$ 4(*NNNN*)

University: Rank 1— 5(*PPPNN*), Rank 2— 3(*PPN*) , Rank 3— 4(*PNNN*)

Publication: Yes 5(*PPPNN*) , No 7(*PPPNNNN*)

Recommendation: good 8(*PPPPPNNN*) , normal 4(*PNNN*)

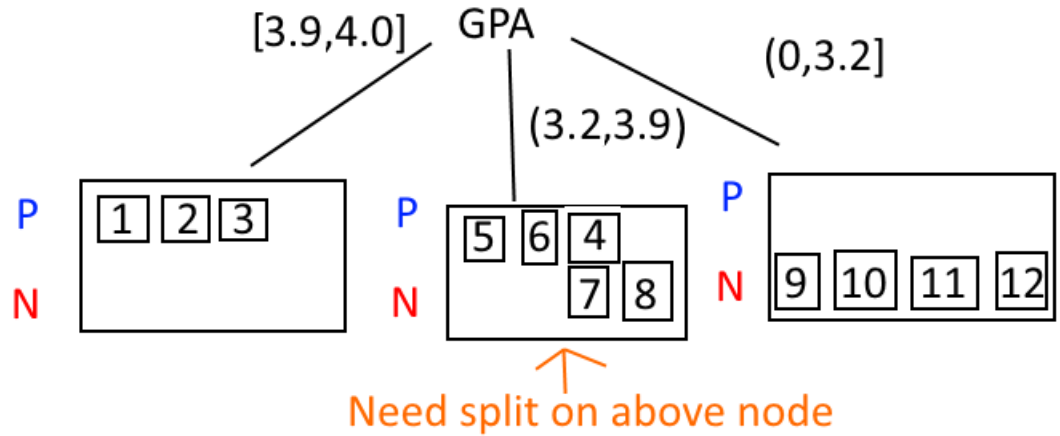
$$\text{Gain(GPA)} = 1 - [\frac{3}{12} B(\frac{3}{3}) + \frac{5}{12} B(\frac{3}{5}) + \frac{4}{12} B(\frac{0}{4})] = 1 - [0.0 + 0.404562747689 + 0.0] = 0.595437252311$$

$$\text{Gain(University)} = 1 - [\frac{5}{12} B(\frac{3}{5}) + \frac{3}{12} B(\frac{2}{3}) + \frac{4}{12} B(\frac{1}{4})] = 1 - [0.404562747689 + 0.229573958514 + 0.270426041486] = 0.095437252395$$

$$\text{Gain(Publication)} = 1 - [\frac{5}{12} B(\frac{3}{5}) + \frac{7}{12} B(\frac{3}{7})] = 1 - [0.404562747689 + 0.574716412687] = 0.020720839624$$

$$\text{Gain}(\text{Recommendation}) = 1 - \left[\frac{8}{12} B\left(\frac{5}{8}\right) + \frac{4}{12} B\left(\frac{1}{4}\right) \right] = 1 - [0.636289335283 + 0.270426041486] = 0.093284623231$$

So we pick GPA as the best Gain attribute in this level



Step II

$$I\left(\frac{2}{5}, \frac{3}{5}\right) = 0.970950594455$$

University: Rank 1— 2(PN), Rank 2— 1(P) , Rank 3— 2(PN)

Publication: Yes 2(PP) , No 3(PNN)

Recommendation: good 5(PPPNN)

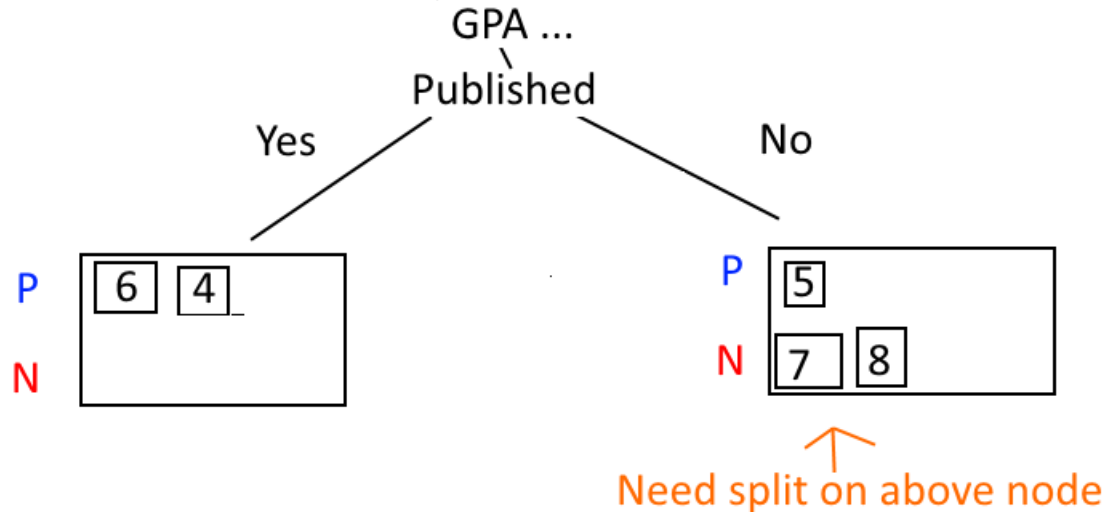
$$\text{Gain}(\text{University}) = 0.970950594455 - \left[\frac{2}{5} B\left(\frac{1}{2}\right) + \frac{1}{5} B\left(\frac{1}{1}\right) + \frac{2}{5} B\left(\frac{1}{2}\right) \right] = 0.970950594455 - [0.4 + 0.0 + 0.4] = 0.170950594455$$

$$\text{Gain}(\text{Publication}) = 0.970950594455 - \left[\frac{3}{5} B\left(\frac{1}{3}\right) + \frac{2}{5} B\left(\frac{2}{2}\right) \right] = 0.970950594455 - [0.550977500433 + 0.0] = 0.419973094022$$

$$\text{Gain}(\text{Recommendation}) = 0.970950594455 - \left[\frac{5}{5} B\left(\frac{3}{5}\right) \right] = 0.970950594455 -$$

$$[0.970950594455+0.0+0.0] = 0$$

So we pick Publication as the best Gain attribute in this level



Step III

$$I(\frac{1}{3}, \frac{2}{3}) = 0.918295834054$$

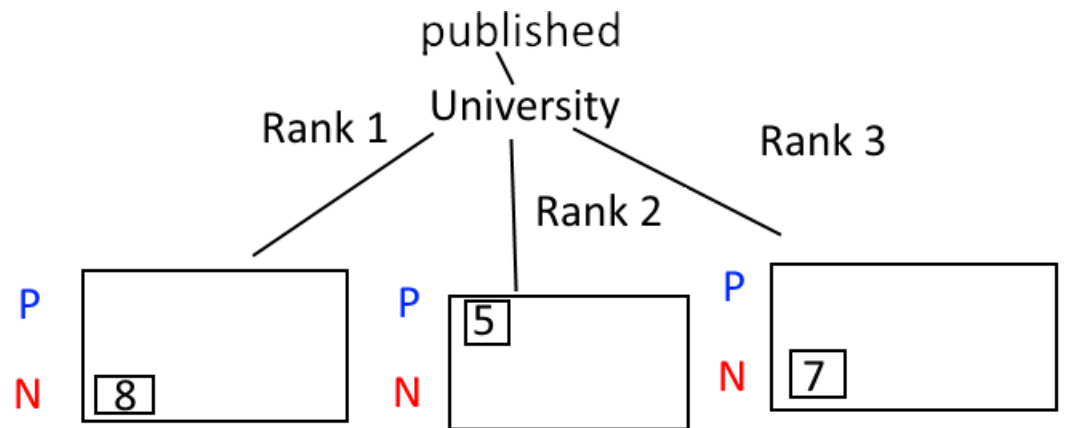
University: Rank 1 — 1(N), Rank 2 — 1(P) , Rank 3 — 1(N)

Recommendation: good 3(PNN)

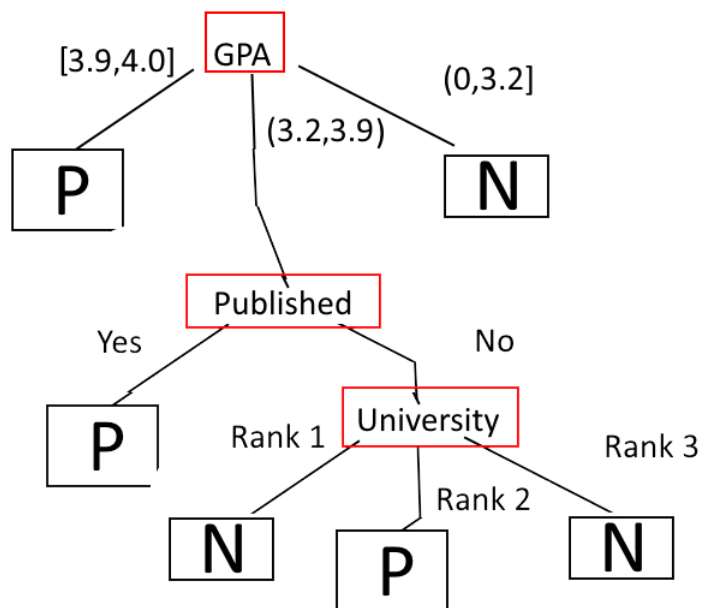
$$\text{Gain(University)} = 0.918295834054 - [\frac{1}{3} B(\frac{0}{1}) + \frac{1}{3} B(\frac{1}{1}) + \frac{1}{3} B(\frac{0}{1})] = 0.918295834054 - [0.0 + 0.0 + 0.0] = 0.918295834054$$

$$\text{Gain(Recommendation)} = 0.918295834054 - [\frac{3}{3} B(\frac{1}{3})] = 0.918295834054 - [0.918295834054] = 0$$

So we pick University as the best Gain attribute in this level



And the final tree to be returned is

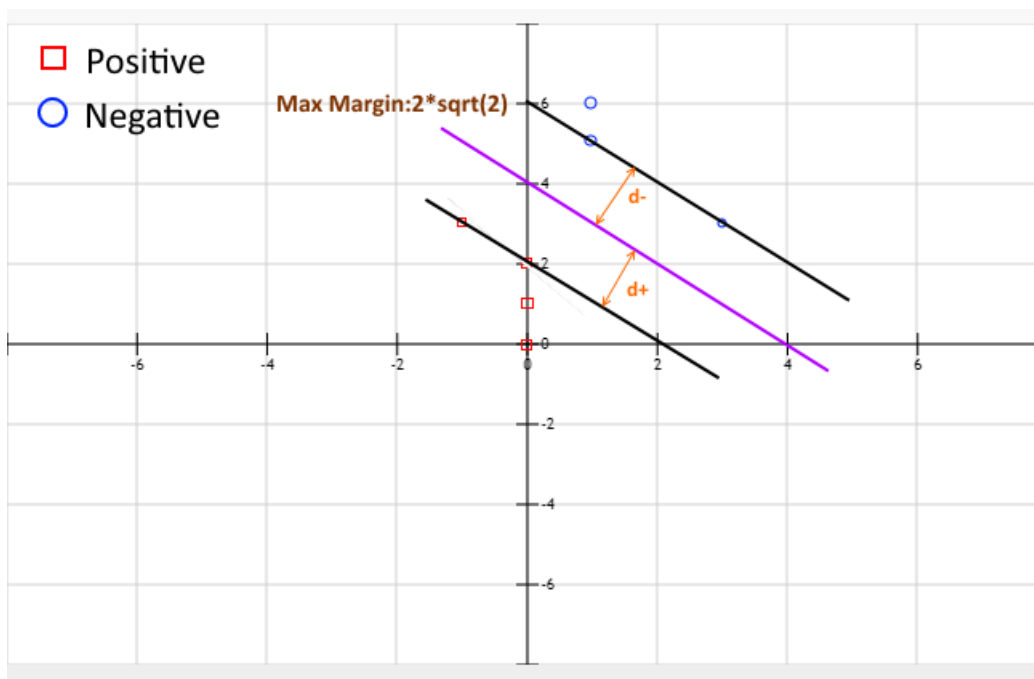


1.3 c.

The decision tree we got from b. is same from the provided one. This is not coincidence, it is believed both tree is generated by applying the decision tree algorithm.

2 problem 3

2.1 a.



2.2 b.

w:(-1,1) b:(4,0)

2.3 c.