Question 2

Part a

1) 4.0 GPA - P, the data match the decision tree
2) 3.9 GPA - P, the data match the decision tree
3) 3.9 GPA - P, the data match the decision tree
4) 3.8 GPA - yes publications - P, the data match the decision tree
5) 3.6 GPA - no publications - rank 2 university - P, the data match the decision tree
6) 3.6 GPA - yes publications - P, the data match the decision tree
7) 3.4 GPA - no publications - rank 3 university - N, the data match the decision tree
8) GPA 3.4 - No publication - Rank 1 University - N, data match the tree
9) GPA 3.2 - N, data match the tree
10) GPA 3.1 - N, data match the tree
11) GPA 3.1 - N, data match the tree
12) GPA 3.0 - N, data match the tree

Part b

For GPA, the information gained is:

$$I(4.0, 3.6, 3.3) = -\frac{1}{4}\log_3\frac{1}{4} - \frac{5}{12}\log_3\frac{5}{12} - \frac{1}{3}\log_3\frac{1}{3} \tag{1}$$

$$I(4.0, 3.6, 3.3) = 0.3155 + 0.3320 + 0.3333 = 0.9808 \tag{2}$$

For university rank, the information gained is:

$$I(rank1, rank2, rank3) = -\frac{5}{12}\log_3\frac{5}{12} - \frac{1}{4}\log_3\frac{1}{4} - \frac{1}{3}\log_3\frac{1}{3} \tag{3}$$

$$I(rank1, rank2, rank3) = 0.3320 + 0.3155 + 0.3333 = 0.9808 \tag{4}$$