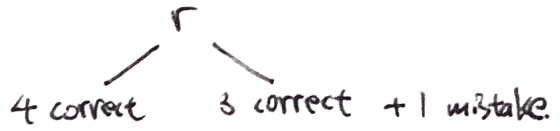
Tianyang Zhang

404-743-024

#1 a). $\pi_i (i \geq 4)$ don't effect value of Y, only π_1, π_2, π_3 effect value of Y.
There are 8 combinations of π_1, π_2, π_3, and there is one mistake among these 8 options.
Therefore, over 2^n training examples, there should be $\frac{2^n}{8}$ mistakes.

b). There is no such split that reduces the number of mistakes by at least one.

Split on π_i, $i \geq 4$ makes no different since it's always $\frac{1}{8}$ mistake.
Split on π_1, π_2, or π_3 would be:



c). $H[Y] = -\frac{1}{8}\log(\frac{1}{8}) - \frac{7}{8}\log(\frac{7}{8}) = 0.543$

d). Do the split in b) on π_1, π_2, π_3, the resulting entropy would be decreased:

$$H[Y|\pi_1] = \left(-\frac{1}{4}\log(\frac{1}{4}) - \frac{3}{4}\log(\frac{3}{4}) - 0\right) \times \frac{1}{2}$$

$$= 0.406$$