

2. Suboptimality IP3.

consider following training set where $X = \{0, 1\}^3$ and $Y = \{0, 1\}$

$$((1, 1, 1), 1)$$

$$((1, 0, 0), 1)$$

$$((1, 1, 0), 0)$$

$$((0, 0, 1), 0)$$

build a decision tree of depth 2.

we denote by H the binary entropy

a) The algorithm first picks the root node, by searching for a feature which maximize information gain.

IG for feature 1 (if we choose $x_{1,0}$ as the root)

$$IG(x_1) = H\left(\frac{1}{2}\right) - \left(\frac{2}{2} H\left(\frac{1}{2}\right) + \frac{1}{2} H(0)\right) = 0$$

$$IG(x_2) = H\left(\frac{1}{2}\right) - \left(\frac{1}{2} H\left(\frac{1}{2}\right) + \frac{1}{2} H\left(\frac{1}{2}\right)\right) = 0$$

$$IG(x_3) = \dots = 0$$

so algorithm pick x_1 as the root.

But three example $((1, 1, 0), 0), ((1, 1, 1), 1), ((1, 0, 0), 1)$

go down one subtree and it doesn't matter what question will ask now, and we can't classify 3 example.

For instance, if next question $x_{1,0}$

either $((1, 1, 0), 0)$ or $((1, 1, 1), 1)$ will mislabeled

and in any case at least one example mislabeled

since we have 4 example so training error is at least $\frac{1}{4}$

b Here is one such tree

