Machine Learning

HW 4 – Decision Trees

Ali Ural

Spring 2025

1. Consider the following set of training examples for an unknown target function $(x_1, x_2) \to y$:

y	x_1	x_2	Count
$+$	T	T	3
$+$	T	F	4
$+$	F	T	4
$+$	F	F	1
$-$	T	T	0
$-$	T	F	1
$-$	F	T	3
$-$	F	F	5

(a) What is the sample entropy for the class label overall, $H(Y)$ from this training data (using log base 2) (3pts)?
The class-label distribution is $P(+) = \frac{12}{21}$ and $P(-) = \frac{9}{21}$. So

$$H(Y) = -\frac{12}{21}\log_2\frac{12}{21} - \frac{9}{21}\log_2\frac{9}{21} \approx 0.985 \text{ bits.}$$

(b) What are the weighed average entropies for branching on variables x_1 and x_2 (6pts)?

x_1

- $x_1 = \text{T}$: 7 positive, 1 negative
  $H = -\frac{7}{8}\log_2\frac{7}{8} - \frac{1}{8}\log_2\frac{1}{8} \approx 0.544.$
- $x_1 = \text{F}$: 5 positive, 8 negative
  $H = -\frac{5}{13}\log_2\frac{5}{13} - \frac{8}{13}\log_2\frac{8}{13} \approx 0.961.$

Weighted entropy:

$$H(Y \mid x_1) = \frac{8}{21}(0.544) + \frac{13}{21}(0.961) \approx 0.802.$$

x_2

- $x_2 = \text{T}$: 7 positive, 3 negative
  $H = -\frac{7}{10}\log_2\frac{7}{10} - \frac{3}{10}\log_2\frac{3}{10} \approx 0.881.$
- $x_2 = \text{F}$: 5 positive, 6 negative
  $H = -\frac{5}{11}\log_2\frac{5}{11} - \frac{6}{11}\log_2\frac{6}{11} \approx 0.994.$

Weighted entropy:

$$H(Y \mid x_2) = \frac{10}{21}(0.881) + \frac{11}{21}(0.994) \approx 0.940.$$

(c) Draw the decision tree that would be learned by the ID3 algorithm without pruning from this training data. If you arrive at a scenario where you have to put a leaf node, but the classes of the data don't all agree, put the probabilities of each class on that leaf node. (6pts)