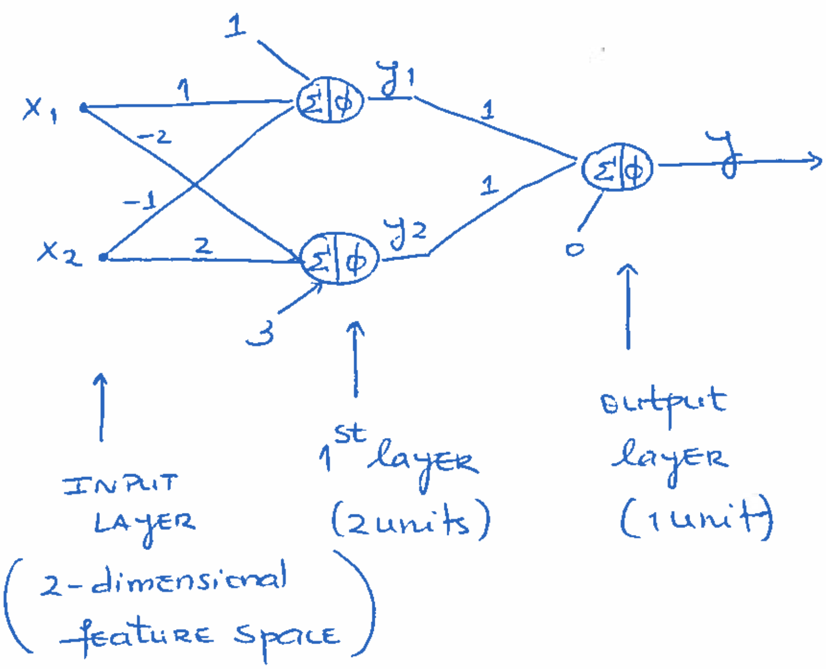
Lecture 15 — Practice Problems

(1)

$$\phi(x) = \begin{cases} 1, & x > 0 \\ -1, & x \leq 0 \end{cases}$$



(1.1)

$$y_1 = \phi(1.x_1 + (-1).x_2 + 1)$$

$$y_2 = \phi(-2.x_1 + 2.x_2 + 3)$$

$$\boxed{y = \phi(1.y_1 + 1.y_2 + 0)}$$

(1.2)

- $(x_1, x_2) = (0,0)$: $y_1 = \phi(1) = 1$
  $y_2 = \phi(3) = 1$
  $y = \phi(2) = 1$ $\Rightarrow$ (0,0) belongs to class 1

- $(x_1, x_2) = (-2, -2.5)$: $y_1 = \phi(1.5) = 1$
  $y_2 = \phi(2) = 1$
  $y = \phi(2) = 1$ $\Rightarrow$ (-2, -2.5) belongs to class 1

- $(x_1, x_2) = (-5, 5)$: $y_1 = \phi(-9) = 0$
  $y_2 = \phi(23) = 1$
  $y = \phi(1) = 1$ $\Rightarrow$ (-5, 5) belongs to class 1

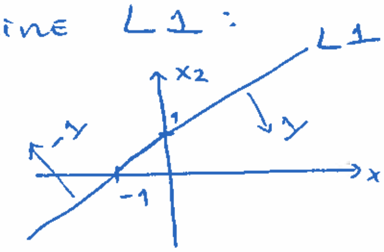
- $(x_1, x_2) = (10, 3)$: $y_1 = \phi(8) = 1$

  $y_2 = \phi(-11) = 0$

  $y = \phi(1) = 1 \Rightarrow (10,3)$ belongs to class 1

(1.3) Each unit in the first layer of a network is "drawing a line", that is, it is partitioning the input space (x_1, x_2) into 2 regions.

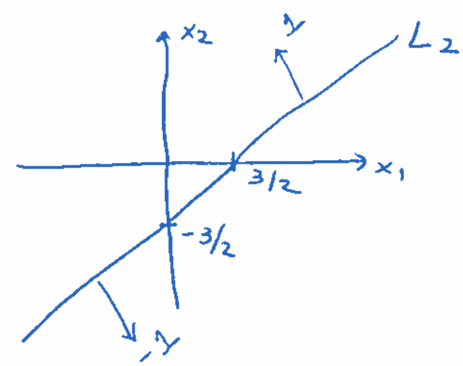
- For unit y_1, we have the line L1:



$$1 . x_1 + (-1) . x_2 + 1 = 0$$

$$\Leftrightarrow \quad x_2 = x_1 + 1$$

- For unit y_2, we have the line L2:

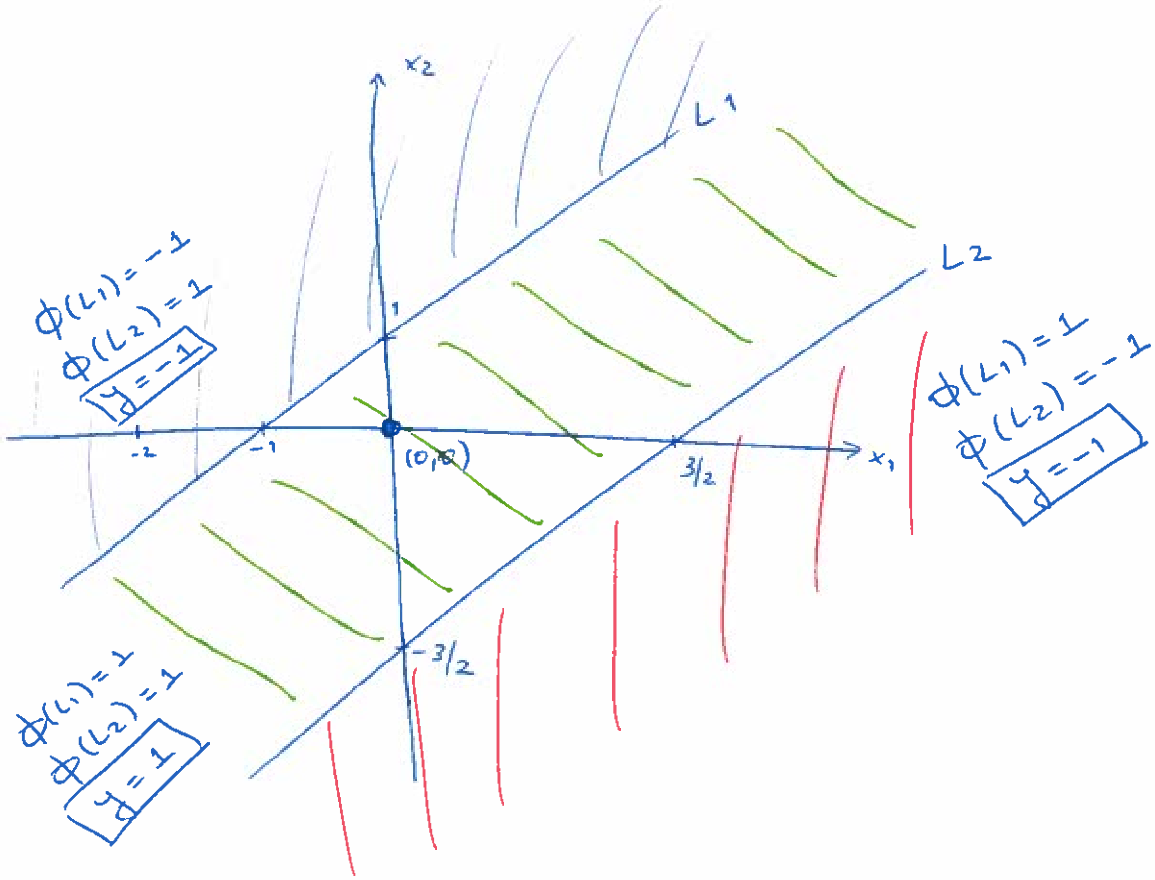


$$-2x_1 + 2x_2 + 3 = 0$$

$$\Leftrightarrow \quad x_2 = x_1 - \frac{3}{2}$$

- Anything above the line L1 will have $L_1 : x_1 - 2x_2 + 1 > 0$ and therefore $\phi(L1) = 1$. Anything below the line L1 will have $L1 < 0$ and therefore $\phi(L1) = -1$.

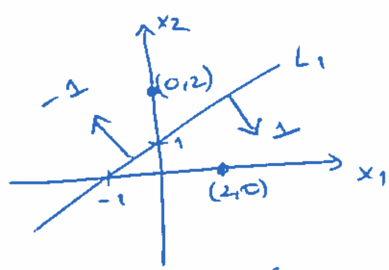
- Similarly, if $L2 > 0 \Rightarrow \phi(L2) = 1$, if $L2 < 0 \Rightarrow \phi(L2) = -1$. Putting it together, we have:



$\Rightarrow$ Any point (x_1, x_2) below L_1 and above L_2 will be labeled as class -1. Any other point will be labeled as class 1.

<u>NOTE</u>: To understand which side of the line will land class 1 vs class -1, you can, ~~for~~ example, choose a point and compute wether the eq. of the line is >0 or <0.

For example,



$L_1 = x_1 - x_2 + 1$

Let's evaluate at point $(x_1, x_2) = (0, 2)$: $L_1 = -1 < 0$
$\Rightarrow \phi(L_1) = -1$

$(x_1, x_2) = (2, 0)$: $L_1 = 3 \Rightarrow \phi(L_1) = 1$

(5

② Decision Tree

x_1: chest pain?	x_2: male?	x_3: smokes?	x_4: exercises?	y: heart attack
1	1	0	1	1
1	1	1	0	1
0	0	1	0	1
0	1	0	1	0
1	0	1	1	1
0	1	1	1	0

Classes:

$y = 1 \Rightarrow$ MEANS "HEART ATTACK"

$y = 0 \Rightarrow$ MEANS "<u>NO</u> HEART ATTACK"

Consider $0 \cdot \log_2(0) = 0$.

① The first step in constructing a decision tree is to compute which feature provides the most information to split the class labels. To do this, we will compute the entropy of each feature x_i. We will choose the feature that minimizes entropy.

- For feature x_1: x_1 is also a binary variable, and so we want to assess the class separability for each binary value of x_1:

FOR $x_1 = 0$:

$$\frac{3}{6}\left(\frac{2}{2}\times\log_2\left(\frac{2}{2}\right) + \frac{1}{4}\times\log_2\left(\frac{1}{4}\right)\right) = -0.25$$

3 0-entries out of 6 possible values

How many ~~of~~ class $y=0$ does $x_1=0$ correctly

How many of class $y=1$ does $x_1=0$ correctly classifies

For $x_1 = 1$: $\frac{3}{6} \times \left(0 \times \log_2(0) + \frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) \simeq -0.156$

So the entropy of x_1 is $H(x_1) = -(-0.25 - 0.156)$
$= 0.406$

- For feature x2: "male?":

$x_2 = 0$: $\frac{2}{6} \times \left(0 \times \log_2(0) + \frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) \simeq -0.167$

$x_2 = 1$: $\frac{4}{6} \times \left(\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) + \frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) \simeq -0.33$

$H(x_2) = -(-0.167 - 0.33) = 0.497$

- For feature x3: "smokes?":

$x_3 = 0$: $\frac{2}{6} \times \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) + \frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) \simeq -0.33$

$x_3 = 1$: $\frac{4}{6} \times \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) + \frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) \simeq -0.54$

$H(x_3) = -(-0.33 - 0.54) = 0.87$

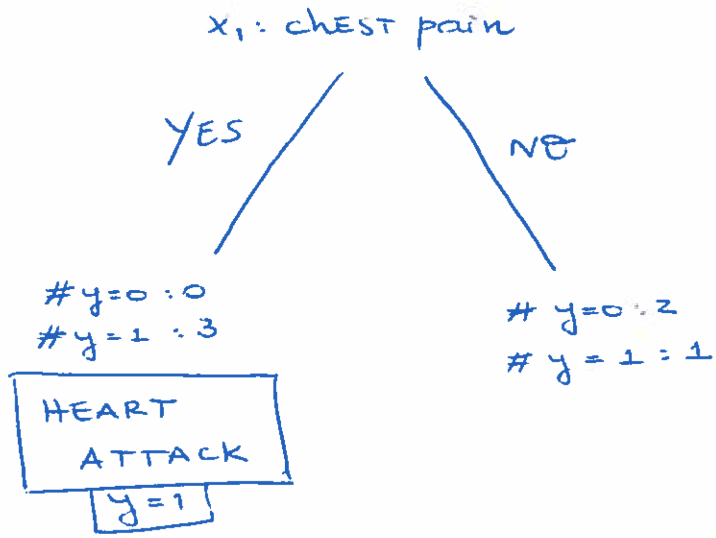
- For feature x_4: "exercises?"

$x_4 = 0$: $\frac{2}{6} \times \left(0 \times \log_2(0) + \frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) \simeq -0.167$

$x_4 = 1$: $\frac{4}{6} \times \left(\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) + \frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) \simeq -0.33$

$H(x_4) = -(-0.167 - 0.33) = 0.497$

So, feature x_1 minimizes entropy and will therefore be the root node.



If $x_1 = 1$ then we will classify that sample with $y = 1$ (HEART ATTACK).

If $x_1 = 0$ then we need to further split the tree.

At this stage, our samples for use are:

x_1: chest pain?	x_2: male?	x_3: smokes?	x_4: exercises?	y: H.A.?
0	0	1	0	1
0	1	0	1	0
0	1	1	1	0

By looking at the table we already see that either feature X_2 or feature x_4 will be able to separate class $y=0$ from $y=1$. But let's compute the entropy:

- For feature x_2:

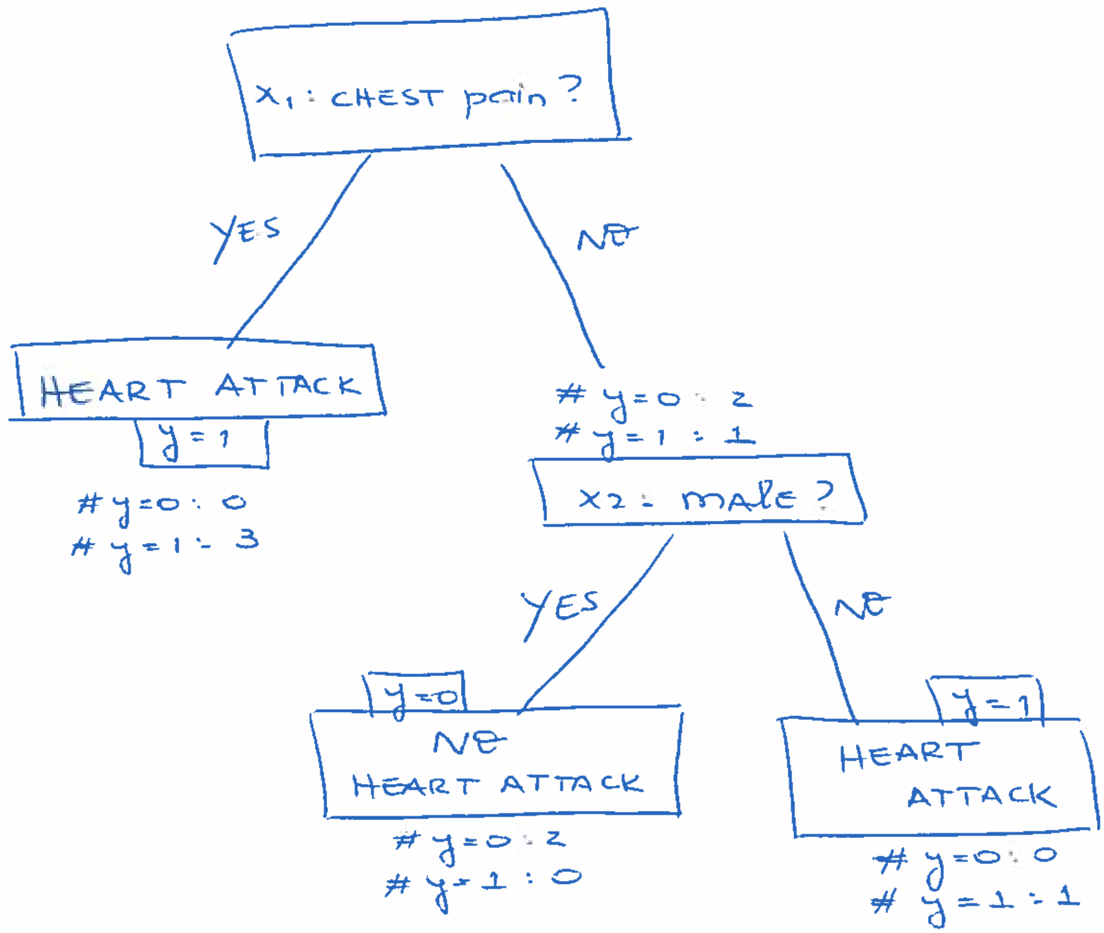
$$x_2 = 0 : \frac{1}{3} \times \left(0 \times \log_2(0) + \frac{1}{1} \times \log_2 \frac{1}{1}\right) = 0$$

$$x_2 = 1 : \frac{2}{3} \times \left(\frac{2}{2} \times \log_2\left(\frac{2}{2}\right) + 0 \times \log_2(0)\right) = 0$$

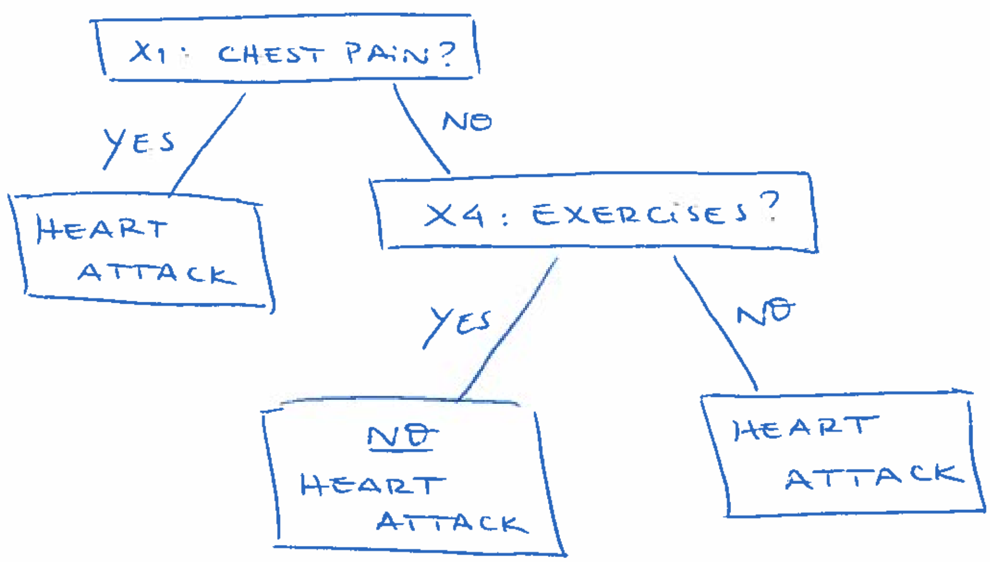
$H(X_2) = 0$

This is already the smallest entropy value

So the final tree is:



An alternative tree is:



↳ In cases when there are ties, we can: (1) flip a coin, ~~and~~ or (2) use cross-validation.