# AI HW 3

**NAME:** 杨睦圳 **ID:** 11813010

## Q1

For the first split, we have:

For $A_1$:

$$\frac{1}{5}(-\log_2 1)+\frac{4}{5}(-\frac{2}{4}\log_2(\frac{2}{4})-\frac{2}{4}\log_2(\frac{2}{4}))=0.8$$

For $A_2$:

$$\frac{2}{5}(-\log_2 1)+\frac{3}{5}(-\frac{1}{3}\log_2(\frac{1}{3})-\frac{2}{3}\log_2(\frac{2}{3}))=0.55$$

For $A_3$:

$$\frac{3}{5}(-\frac{2}{3}\log_2(\frac{2}{3})-\frac{1}{3}\log_2(\frac{1}{3}))+\frac{2}{5}(-\frac{1}{2}\log_2(\frac{1}{2})-\frac{1}{2}\log_2(\frac{1}{2}))=0.95$$

Since 0.95 > 0.8 > 0.55, choose $A_2$ for the first split.

For the second split, we only need to consider the remaining ($x_3$, $x_4$, $x_5$):

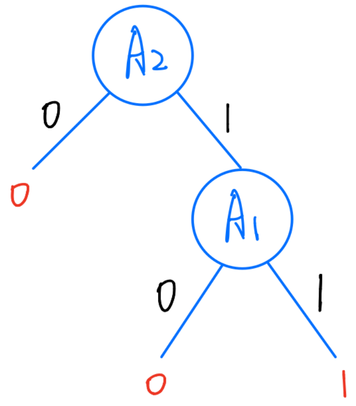
For $A_1$ :

$$\frac{1}{3}(-\log_2 1)+\frac{2}{3}(-\log_2 1)=0$$

For $A_3$:

$$\frac{2}{3}(-\frac{1}{2}\log_2(\frac{1}{2})-\frac{1}{2}\log_2(\frac{1}{2}))+\frac{1}{3}(-\log_2 1)=0.67$$

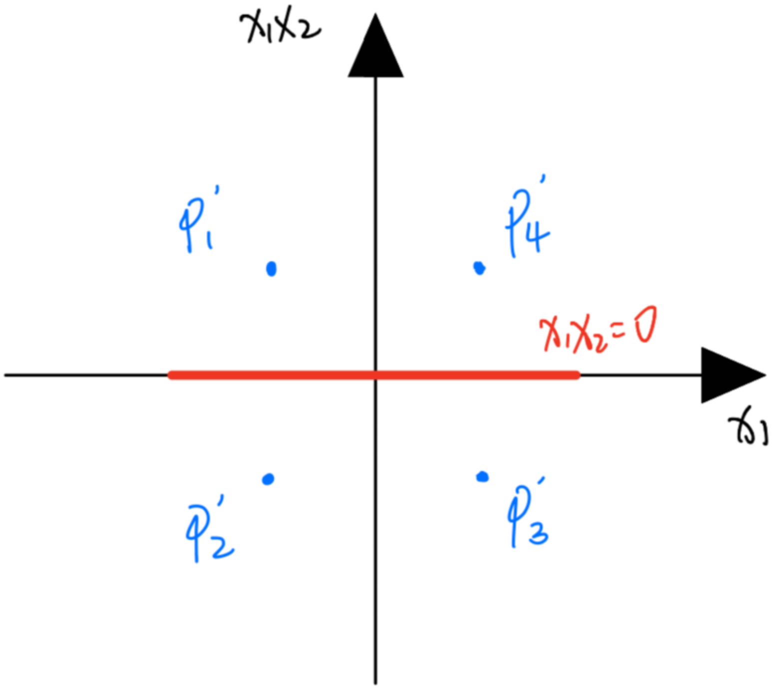
Since 0.67 > 0, choose $A_1$ for the second split.



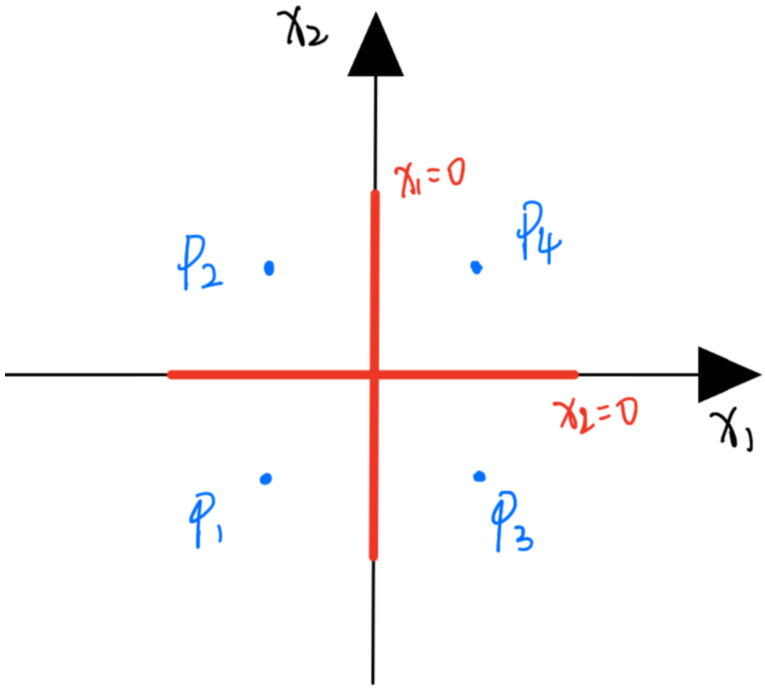
## Q2

| i | Point before map $P_i$ | Mapped point $P_i'$ | Function value |
|---|---|---|---|
| 1 | [-1, -1] | [-1, 1] | -1 |
| 2 | [-1, 1] | [-1, -1] | 1 |
| 3 | [1, -1] | [1, -1] | 1 |
| 4 | [1, 1] | [1, 1] | -1 |

The line $x_1x_2 = 0$ is the maximal margin separator with the margin of 1.




It is corresponds to the lines $x_1 = 0$ and $x_2 = 0$ in the origin space.



# Q3.

## a.

Assume that the activation function at each node is $g(x) = a * x + b$, the the j<sup>th</sup> unit $a_j = \sum_i \omega_{i,j} x_i$

The output of the hidden layer are $H_j = g(a_j) = a \sum_i \omega_{i,j} x_i + b$

The output of the output layer are

$$g(\sum_j \omega_{j,k} H_j) = a * (\sum_j \omega_{j,k} (a \sum_i \omega_{i,j} x_i + b)) + b = a^2 \sum_j \omega_{j,k} \sum_i \omega_{i,j} x_i + b(c * \sum_j \omega_{j,k} + 1)$$

Since $a^2 \sum_j \omega_{j,k} \sum_i \omega_{i,j}$ and $b(c * \sum_j \omega_{j,k} + 1)$ are constant, let them to be $k, d$, then there is a network with no hidden units that computes the same function with the activation function $g(x) = kx + d$

## b.

The method in (a) can be used to reduce an n-layer network to (n-1)-layer network. Then inductively, we can reduce it to (n-2)-layer network, (n-3)-layer network, ..., single-layer network.

## c.

For the network with one hidden layer and linear activation functions has input and output $n$ nodes and $h$ hidden nodes, we can know that it has $nh + nh = 2nh$ weights. Without the hidden layer, there will be $n^2$ weights. When $h << n,\ 2hn < n^2$, so there will be less weights in the network with hidden layer.