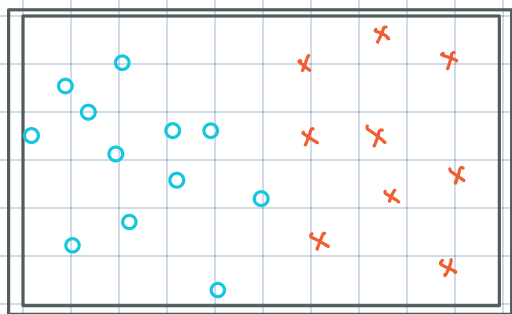


▷ Classification

$$\text{Data} = \{(\vec{x}_i, C_i)\}, C_i \in \{0, 1\}$$



為什麼比較好?

① **Method 1**: Find a function: $\mathbb{R}^2 \rightarrow \mathbb{R}$
s.t., $H(\vec{x}_i) = C_i, C_i \in \{0, 1\}$

② **Method 2**: One-time encoding (丟到高維度向量)

$$C_i \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

Find the function, $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ s.t., $H(\vec{x}_i) = C_i$

$$H(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \end{bmatrix}$$

<來說明為什麼 Method 2 比 Method 1 來的好>

- ① In classification problem, if we use method 1, $C_i \in \{0, 1\}$. Although this definition is simple but it implied a numerical value order "0 < 1". For the classification problem, any class neither have size relationship nor distance relationship. Thus it may be mistaken by the mathematic.
 - ② One-time encoding let C_i to express 2-dimension one-hot vector (i.e., $C_i \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$) s.t. the label in two orthogonal vector of \mathbb{R}^2 , avoiding the fake order of the numerical value.
 - ③ Moreover, Using Method 2, we can also put the cross-entropy loss in our training. If the model let the correct class probability higher, then the loss function is low, the model is overly confident about the wrong class, the loss must be very large.
 - ④ We can image } The Method 1 put A, B in two endpoint of the line, like $B > A$
 } The Method 2 put A, B in coordinate system and different axis, which is independent.
- Hence, if we solve a classification problem, we use Method 2 !!