

機器學習基石(林軒田老師)筆記

Lecture 2: Learning to Answer Yes/No

About threshold in PLA

1. Why use threshold?

The following is page 4 of 02_handout of Machine Learning Foundations.

Learning to Answer Yes/No

Perceptron Hypothesis Set

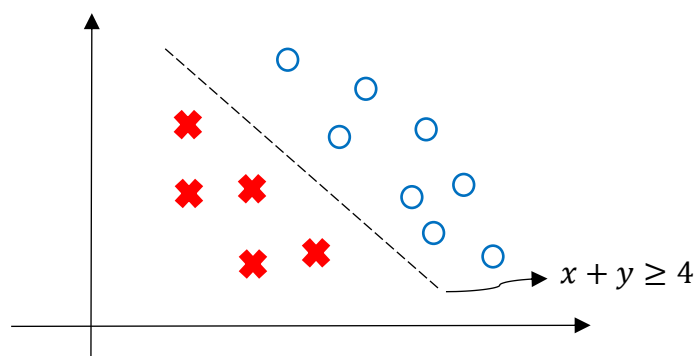
Vector Form of Perceptron Hypothesis

$$\begin{aligned}h(\mathbf{x}) &= \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) - \text{threshold} \right) \\&= \text{sign} \left(\left(\sum_{i=1}^d w_i x_i \right) + \underbrace{(-\text{threshold})}_{w_0} \cdot \underbrace{(+1)}_{x_0} \right) \\&= \text{sign} \left(\sum_{i=0}^d w_i x_i \right) \\&= \text{sign} (\mathbf{w}^T \mathbf{x})\end{aligned}$$

- each 'tall' \mathbf{w} represents a hypothesis h & is multiplied with 'tall' \mathbf{x} —will use tall versions to simplify notation

what do perceptrons h 'look like'?

PLA find a hyperplane to separate data with different values(-1 and 1), but the hyperplane which separate data usually doesn't contain origin.



We need to add one dimension to the data. If data is of dimension two, we want to find a good $h(\mathbf{x}) = \text{sign}(w_0 + w_1 x_1 + w_2 x_2)$. In the example above, $x + y \geq 4$ will separate the data perfectly.

$$x + y \geq 4 \Leftrightarrow x + y - 4 \geq 0 \leftrightarrow h(x) = \text{sign}(-4 + x + y) \quad (a)$$

Now $x + y \geq 4$ corresponds to $(-4, 1, 1)$ a 3-dimensional vector. And $(-4, 1, 1)$ also correspond to a plane $-4p + q + r = 0$ in 3-dimensional space which contains the origin.

$$h(x) = \text{sign}(-4 + x + y) \leftrightarrow (-4, 1, 1) \leftrightarrow -4p + q + r = 0 \quad (b)$$

Combine (a) and (b), we have:

$$x + y \geq 4 \leftrightarrow -4p + q + r = 0$$

That is, a line in 2-dimensional space (not necessary contains the origin) corresponds to a plane (contains the origin) in 3-dimensional space.

Note that $-4p + q + r = 0 \leftrightarrow x + y \geq 4$, while $4p - q - r = 0 \leftrightarrow x + y \leq 4$ in this correspondence.

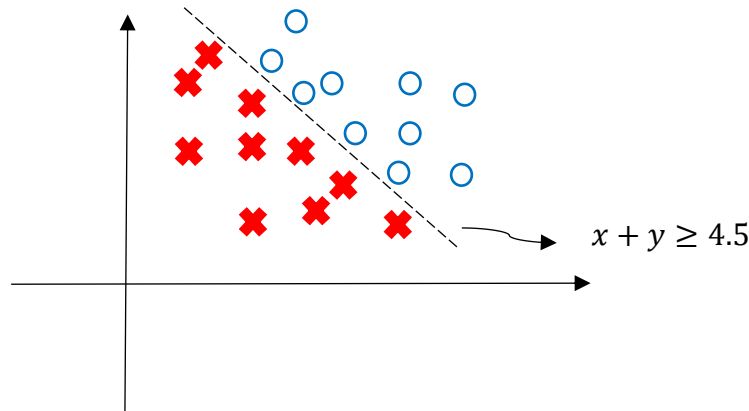
Therefore, to find a good $h(x)$ is equal to find a good hyperplane.

2. Integral coefficient $w_{0,0}$?

In PLA, we have:

$$W_{t+1} \leftarrow W_t + y_{n(t)} X_{n(t)}$$

We usually let $W_0 = (w_{0,0}, w_{0,1}, \dots, w_{0,d}) = (0, 0, \dots, 0)$, and we have $x_{n(t),0} = 1$ for all data. Thus we have $w_{t,0} \in \mathbb{Z}$ for all t . What if we have data like this:



What $w_{t,0}$ will we get when PLA stop?

In fact, we will get some W_t close to $(-9, 2, 2)$ since

$$x + y \geq 4.5 \Leftrightarrow 2x + 2y \geq 9$$

And by properties of real numbers, we can find a good W_t even if all $w_{t,0} \in \mathbb{Z}$.