P1: $\hat{t} = \text{argmax}\ \hat{p}(t|x) = \text{argmax}\ \hat{p}(t)\,p(x|t)$

∴ linear.

P2:

	0,0	0,1	1,0	1,1
C_1	+	+	+	−
C_2	−	+	+	+

set: $(-1,-1,1,1,\ 1,1,\ 1.5,-0.5,-1)$

P3: $h = \dfrac{1}{1+e^{-(w_1^T x + b_1)}}$

$y = \dfrac{1}{1+e^{-(w_2 h + b_2)}}$

$y = f(x) = \sigma(w_2\,\sigma(w_1^T x + b_1) + b_2)$.

y of this 2-layer classifier is a linear combination of x.

cannot achieve none-linearity.

P4: $f(z) = z \qquad f'(z) = 1$.

$f(z) = \sigma(z) = \dfrac{1}{1+e^{-z}} \qquad f'(z) = -\dfrac{1}{(1+e^{-z})^2} \times \dfrac{d}{dz}(1+e^{-z}) = \dfrac{e^{-z}}{(1+e^{-z})^2} = z\cdot(1-z)$

$f(z) = \text{ReLU}(z) = \max(0,z) \qquad f'(z) = \begin{cases} 1 & f(z) = z \\ \text{not differentiable} & f(z) = 0 \end{cases}$