$$p(t(x)=+1 \mid x) = p(+1 \mid x) = p(C_1 \mid x) = \sigma(z) = \frac{1}{1+e^{-z}}$$

$$z = w^T x + w_0$$

$$= \frac{1}{1+e^{-w^T x - w_0}}$$

$$p(C_1 \mid x) \geq 0.5 \iff \frac{1}{1+e^{-z}} \geq \frac{1}{2} \iff 1+e^{-z} \leq 2 \iff e^{-z} \leq 1 \;\Big|\ln$$

$$\iff -z \leq 0 \iff z \geq 0.$$

$\vec{t} = [t_1, t_2, \dots, t_N]$

$X = [x_1, x_2, \dots, x_N]$

Likelihood: $p(t_1, t_2, \dots, t_N \mid w, x_1, x_2, \dots, x_N)$

$$= p(\vec{t} \mid w, X)$$

MLE: $\hat{w} = \arg\max_{w} p(t \mid w, X) = \arg\min_{w} -\ln p(t \mid w, X).$

NLL $\rightarrow$ cross-entropy.

$$\hat{w} = \arg\min_{w} -\ln p(\vec{t}\,|w, X) \underset{\uparrow \text{ by iid assumptation}}{=} \arg\min_{w} -\ln \prod_{n=1}^{N} p(t_n|w, x_n)$$

$$= \arg\min_{w} -\sum_{n=1}^{N} \ln \underbrace{p(t_n|w, x_n)}_{= \; h_n^{t_n}(1-h_n)^{1-t_n}}$$

where $h_n = p(C_1|x_n) = \sigma(w^T x_n)$

Case 0: $t_n = 0 \Rightarrow p(t_n = 0|w, x_n) = 1 - p(t_n = 1|w, x_n) = 1 - h_n$

$$h_n^{t_n}(1-h_n)^{1-t_n} = h_n^{0}(1-h_n)^{1-0} = 1 - h_n$$

Case 1: $t_n = 1 \Rightarrow p(t_n = 1|w, x_n) = h_n$

$$h_n^{t_n}(1-h_n)^{1-t_n} = h_n^{1}(1-h_n)^{0} = h_n$$

· use gradient descent in HW4

$$= \boxed{\arg\min_{w} -\sum_{n=1}^{N} t_n \ln h_n + (1-t_n)\ln(1-h_n)}$$