Linear Discrimination

$$P(y=1 \mid x) = \delta$$

$$P(y=2 \mid x) = 1-\delta$$

Choose C#1 if $\begin{cases} \delta > 0.5 \\ \delta/1-\delta > 1 \\ \log[\delta/(1-\delta)] > 0 \end{cases}$

$$\log\left[\frac{P(y=1|x)}{P(y=2|x)}\right] = \log\left[\frac{p(x|y=1)}{p(x|y=2)}\right] + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

$p(x|y=1) \to N(x;\mu_1,\Sigma)$

$p(x|y=2) \to N(x;\mu_2,\Sigma)$

$$\Sigma_1 = \Sigma_2 = \Sigma$$

$$N(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^T \cdot \Sigma^{-1} \cdot (x-\mu)\right]$$

$$\log\left[\frac{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}} \exp\left[-\frac{1}{2}(x-\mu_1)^T \cdot \Sigma^{-1} \cdot (x-\mu_1)\right]}{\cancel{(2\pi)^{-D/2}} \cdot \cancel{|\Sigma|^{-1/2}} \exp\left[-\frac{1}{2}(x-\mu_2)^T \cdot \Sigma^{-1} (x-\mu_2)\right]}\right] + \log\left[\frac{P(y=1)}{P(y=2)}\right]$$

$$\boxed{\frac{\exp(a)}{\exp(b)} = \exp(a-b)}$$

$$= \underbrace{\left[\Sigma^{-1}(\mu_1-\mu_2)\right]^T}_{w} \cdot x + \underbrace{\left[-\frac{1}{2}(\mu_1+\mu_2)^T \Sigma^{-1}(\mu_1-\mu_2) + \log\left[\frac{P(y=1)}{P(y=2)}\right]\right]}_{w_0}$$

$$= w^T \cdot x + w_0$$

$$\hat{w} = \hat{\Sigma}^{-1} \cdot (\hat{\mu}_1 - \hat{\mu}_2)$$

$\hat{\Sigma}$ → sample covariance of all data points

$\hat{\mu}_2$ → sample mean of second class

$\hat{\mu}_1$ → sample mean of first class

$$\hat{w}_0 = -\frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)^T \hat{\Sigma}^{-1} \cdot (\hat{\mu}_1 - \hat{\mu}_2) + \log\left[\frac{\hat{p}(y=1)}{\hat{p}(y=2)}\right]$$

$\hat{p}(y=1)$ → frequency of first class.

$\hat{p}(y=2)$ → frequency of second class

$$\exp\left[\log\left[\frac{\delta}{1-\delta}\right]\right] = \exp\left[w^T \cdot x + w_0\right]$$

$$\frac{\delta}{1-\delta} = \exp\left[w^T \cdot x + w_0\right] \Rightarrow \delta = \exp\left[w^T \cdot x + w_0\right] - \delta \exp\left[w^T \cdot x + w_0\right]$$

$$\delta\left(1 + \exp\left[w^T \cdot x + w_0\right]\right) = \exp\left[w^T \cdot x + w_0\right]$$

$$\delta = \frac{\exp\left[w^T \cdot x + w_0\right]}{1 + \exp\left[w^T \cdot x + w_0\right]}$$

a) if $w^T.x + w_0 > 0 \Rightarrow \delta > 0.5$

b) if $w^T.x + w_0 = 0 \Rightarrow \delta = 0.5$

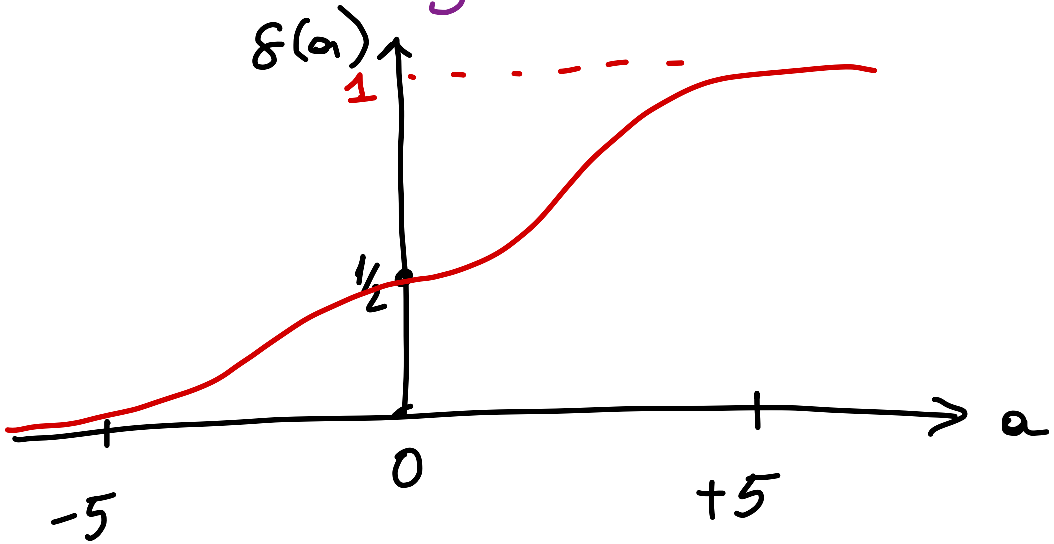
c) if $w^T.x + w_0 < 0 \Rightarrow \delta < 0.5$

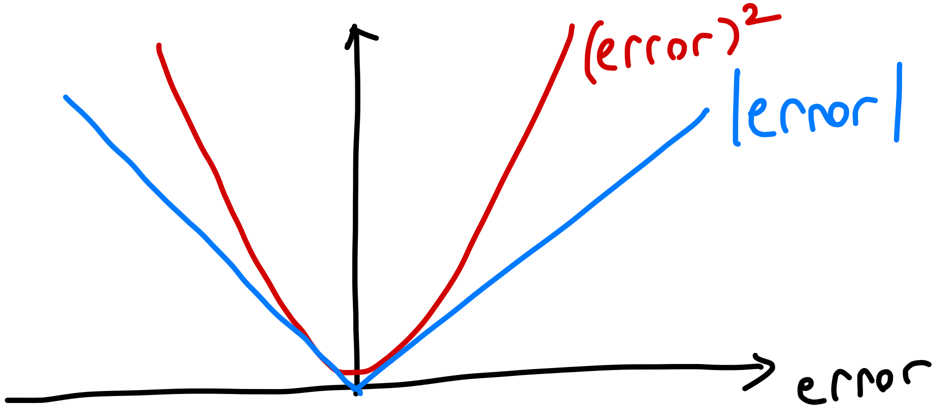
$$\delta = \frac{\exp(w^T.x+w_0) \,/ \exp(w^T.x+w_0)}{[1+\exp(w^T.x+w_0)] \,/ \exp(w^T.x+w_0)} = \underbrace{\frac{1}{1+\exp[-(w^T.x+w_0)]}}_{\text{sigmoid function}}$$

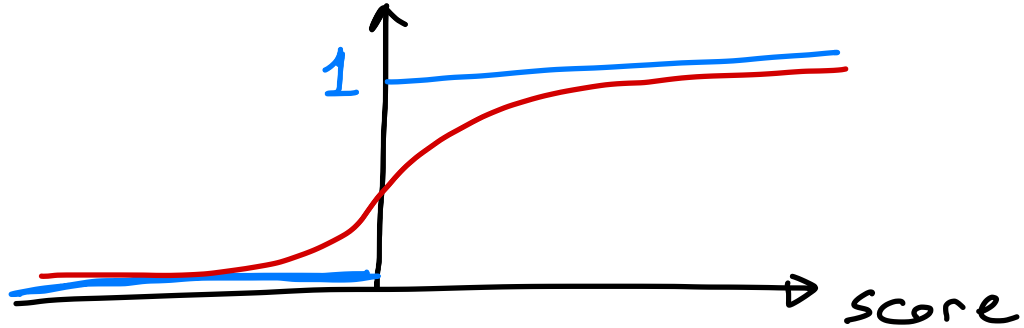
$$\delta(a) = \frac{1}{1+\exp(-a)}$$

$$\delta(-5) = \frac{1}{1+\exp(5)} \cong 0$$

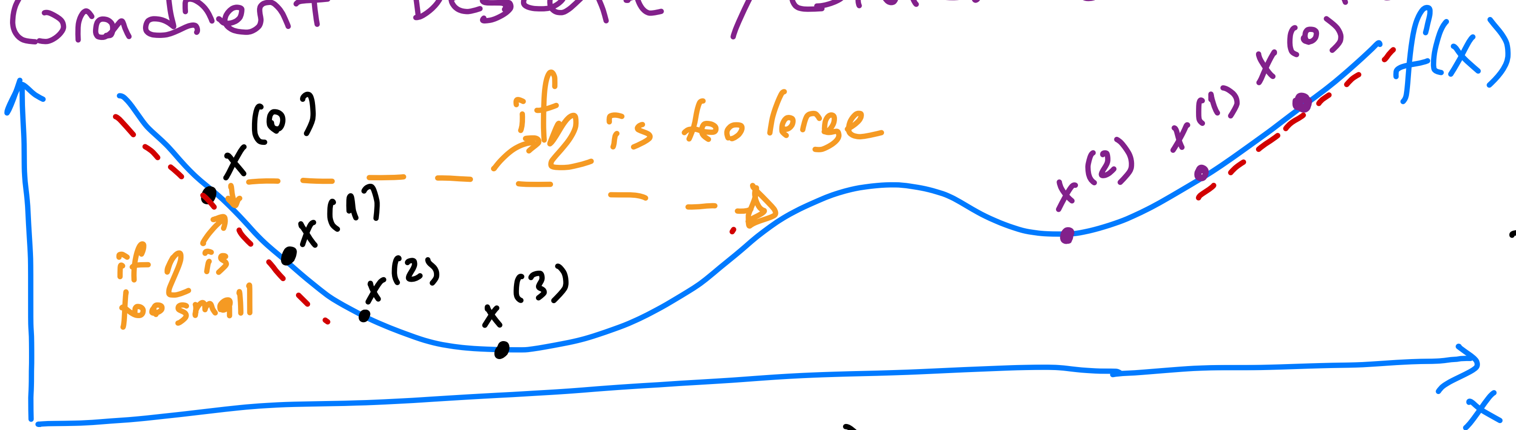
$$\delta(+5) = \frac{1}{1+\exp(-5)} \cong 1$$







Gradient Descent / Gradient Ascent



$$\frac{\delta f(x)}{\delta (x)} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$x^* = \arg\min_{x} f(x)$$

$$\underbrace{\Delta x}_{\substack{\text{step} \\ \text{(update)}}} = - \underbrace{\eta}_{\substack{\text{step} \\ \text{size}}} \cdot \underbrace{\frac{\partial f(x)}{\partial x}}_{\substack{\text{derivative} \\ \text{(slope)}}}$$

$$x^{(t+1)} = x^{(t)} + \Delta x$$

$$= x^{(t)} - \overset{?}{\eta} \cdot \frac{\partial f(x)}{\partial x}$$

$$(w^*, w_0^*) = \arg \min_{(w, w_0)} E[w, w_0 \mid \underbrace{\mathcal{X}}_{\text{training set}}]$$

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad y_i \in \{0, 1\} \qquad x_i \in \mathbb{R}^D$$

$1 \rightarrow$ positive class (success)

$0 \rightarrow$ negative class (failure)

$$y_i \mid x_i \sim \text{Bernoulli}(y_i ; \underbrace{\hat{p}(y = 1 \mid x_i)}_{\text{success probability } (\hat{y}_i)})$$

$\delta = \hat{y}_i$ = probability of x_i being from the positive (+) class

$$\boxed{\text{BERNOULLI} \quad p(x) = p^x (1-p)^{1-x}}$$

$p \rightarrow$ success probability

$$\text{likelihood}(w, w_0 \mid \mathcal{X}) = \prod_{i=1}^{N} \left[[\hat{y}_i]^{y_i} [1 - \hat{y}_i]^{1-y_i} \right]$$

$$\text{loglikelihood}(w, w_0 \mid \mathcal{X}) = \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]$$

$$E[w, w_0 \mid \mathcal{X}] = \underbrace{- \sum_{i=1}^{N} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right]}_{\text{negative loglikelihood}}$$

$$\text{minimize } -\sum_{i=1}^{N}\left[y_i \underline{\log(\hat{y}_i)} + (1-y_i)\underline{\log(1-\hat{y}_i)}\right]$$

with respect to: $[w, w_0]$

$$\frac{1}{1+\exp[-[w^T.x_i+w_0]]} = \hat{y}_i$$

$$\frac{\partial Error}{\partial w} = ? \qquad \frac{\partial Error}{\partial w_0} = ?$$

$$w^{(t+1)} = w^{(t)} - \eta \cdot \frac{\partial Error}{\partial w}$$

$$w_0^{(t+1)} = w_0^{(t)} - \eta \frac{\partial Error}{\partial w_0}$$

$$\text{sigmoid}(a) = \frac{1}{1+\exp(-a)} \quad \begin{matrix} \rightarrow f(a) \\ \rightarrow g(a) \end{matrix}$$

$$\boxed{\begin{gathered} \frac{\partial \log(a)}{\partial a} = \frac{1}{a} \\ \frac{\partial f(a)/g(a)}{\partial a} = \frac{\frac{\partial f(a)}{\partial a} g(a) - f(a)\frac{\partial g(a)}{\partial a}}{[g(a)]^2} \end{gathered}}$$

Exercise

$$\frac{\partial\, \text{sigmoid}(a)}{\partial a} = \text{sigmoid}(a).[1-\text{sigmoid}(a)]$$

Hint:

$$\frac{\partial\, \text{sigmoid}(a)}{\partial a} = \frac{0.[1+\exp(-a)] - 1.\frac{\partial(1+\exp(-a))}{\partial a}}{[1+\exp(-a)]^2}$$

$$\log[\hat{y}_i] = \log\Big[\underbrace{\text{sigmoid}(\underbrace{w^T \cdot x_i + w_0}_{c})}_{d}\Big]$$

$$\frac{\partial \log[\hat{y}_i]}{\partial w} = \underbrace{\frac{\partial \log[\hat{y}_i]}{\partial d}}_{1/d} \cdot \underbrace{\frac{\partial d}{\partial c}}_{d(1-d)} \underbrace{\frac{\partial c}{\partial w}}_{x_i}$$

[previous page] [previous page]

$$\frac{\partial \log[1-\hat{y}_i]}{\partial w} = ?$$

Exercise

$$\boxed{\frac{\partial \text{Error}}{\partial w} = -\sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot x_i \qquad \frac{\partial \text{Error}}{\partial w_0} = -\sum_{i=1}^{N} (y_i - \hat{y}_i)}$$

$$\Delta w = -\eta \frac{\partial \text{Error}}{\partial w} = \eta \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot x_i$$

$$\Delta w_0 = -\eta \frac{\partial \text{Error}}{\partial w_0} = \eta \sum_{i=1}^{N} (y_i - \hat{y}_i)$$

STEP #1: Initialize w, w_0 and decide η.

initialize them to very small values
for example Uniform $[-0.001, +0.001]$

STEP #2: Calculate Δw and Δw_0.

STEP #3: Update w and w_0 using Δw and Δw_0.

$$w^{(t+1)} = w^{(t)} + \Delta w^{(t)}$$
$$w_0^{(t+1)} = w_0^{(t)} + \Delta w_0^{(t)}$$

STEP #4: Go to STEP#2 if there is a change in the parameters [i.e., $\|\Delta w\| \neq 0$, $|\Delta w_0| \neq 0$]

if $\|\Delta w\| < \epsilon$ & $|\Delta w_0| < \epsilon$ where ϵ is a very small number such as 10^{-10} we should stop the algorithm.