Linear Discrimination

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

choose C_1 if $\begin{cases} \delta > 0.5 \\ \delta/(1-\delta) > 1 \\ \log\left[\dfrac{\delta}{1-\delta}\right] > 0 \end{cases}$

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

$$p(x \mid y=1) \rightarrow N(x; \mu_1, \Sigma) \qquad p(x \mid y=2) \rightarrow N(x; \mu_2, \Sigma)$$

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

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

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

$$= \underbrace{\left[\Sigma^{-1}.(\mu_1-\mu_2)\right]^T}_{W}.x + \underbrace{\left[-\frac{1}{2}(\mu_1+\mu_2)^T\Sigma^{-1}(\mu_1-\mu_2)+\log\left[\frac{P(y=1)}{P(y=2)}\right]\right]}_{W_0}$$

$$= W^T.x + W_0$$

sample mean of C_1

sample mean of C_2

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

sample covariance of all data points

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

prior probability estmates for C_1 and C_2

$$\delta = ?$$

$$\exp\left[\log\left[\frac{\delta}{1-\delta}\right]\right] = \exp\left[W^T.x+W_0\right]$$

$$\frac{\delta}{1-\delta} = \exp\left[W^T.x+W_0\right] \Rightarrow \delta = \exp\left[W^T.x+W_0\right] - \delta\exp\left[W^T.x+W_0\right]$$

$$\delta\left(1+\exp\left[W^T.x+W_0\right]\right) = \exp\left[W^T.x+W_0\right]$$

$$\delta = \frac{\exp\left[W^T.x+W_0\right]}{1+\exp\left[W^T.x+W_0\right]}$$

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

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

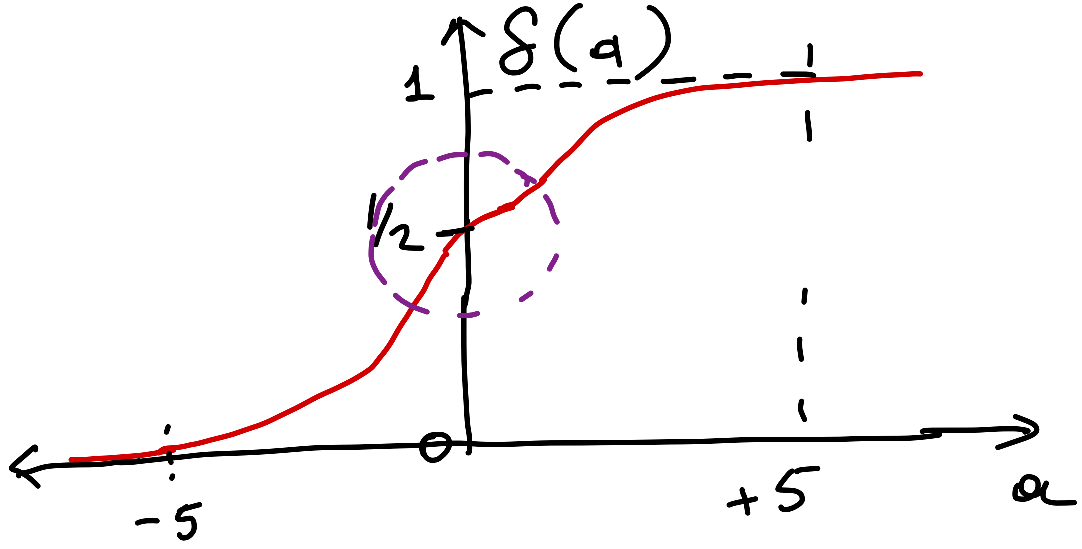
c) if $w^T.x + w_0 < 0 \Rightarrow \delta < 0.5$

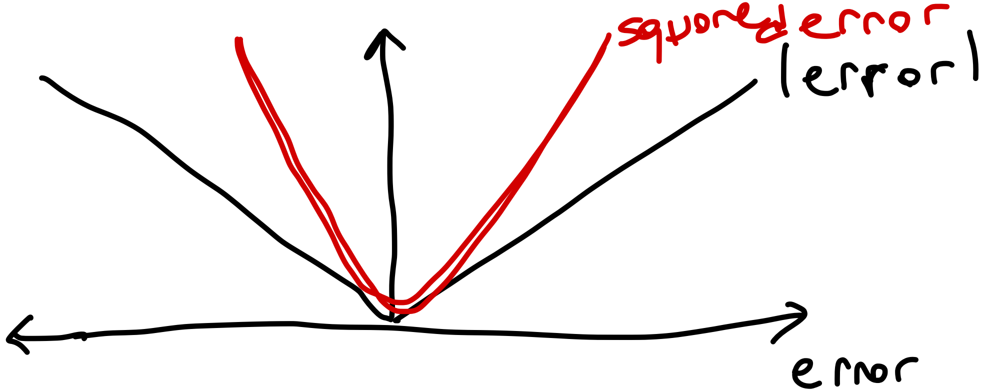
$$\delta = \frac{\exp[\cancel{w^T.x+w_0}] \,/\, \exp(\cancel{w^T.x+w_0})}{[1+\exp[w^T.x+w_0]] \,/\, \exp(w^T.x+w_0)} = \underbrace{\frac{1}{1+\exp[-[w^T.x+w_0]]}}_{\text{sigmoid function}}$$

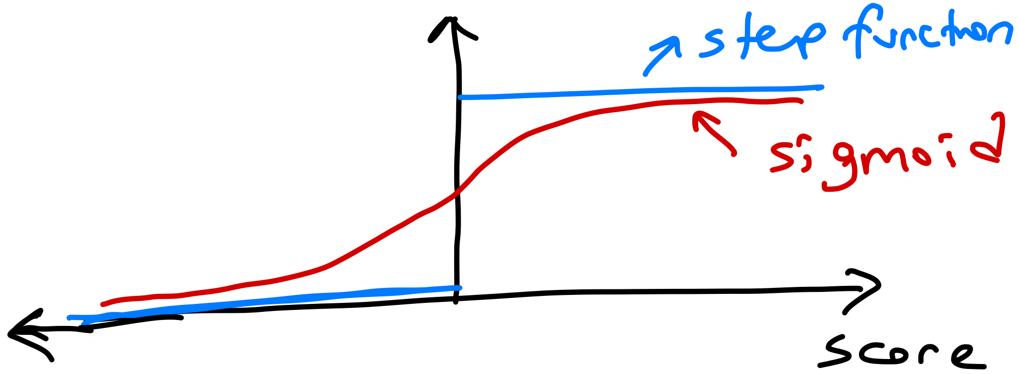
$$\delta(a) = \frac{1}{1+\exp(-a)}$$

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

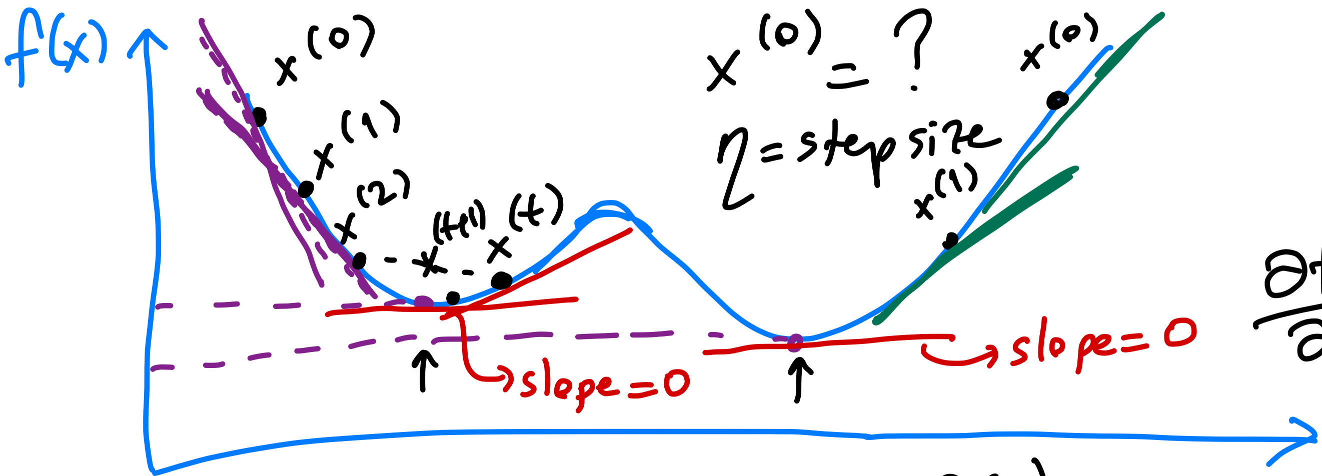
$$\delta(+5) = \frac{1}{1+\exp(-5)} \cong 1$$







Gradient Descent / Gradient Ascent



$$x^* = \arg\min_x f(x)$$

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

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

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

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

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

$0 \to$ negative class

$1 \to$ positive class

Bernoulli distribution

$$p(x) = p^x \cdot (1-p)^{1-x}$$

$p \to$ success probability

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

random variables

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

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

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

$$E[w, w_0 \mid \mathcal{X}] = -\text{log-likelihood}(w, w_0 \mid \mathcal{X})$$

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

with respect to $[w, w_0]$

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

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

$$\frac{f'(a).g(a)-f(a)g'(a)}{[g(a)]^2}$$

Exercise#4

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

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

Hint!

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

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

$$\frac{\partial \log[\hat{y}_i]}{\partial \underline{\underline{w}}} = \underbrace{\frac{\partial \log[\hat{y}_i]}{\partial c}}_{\text{derivative } \log(\cdot)} \; \underbrace{\frac{\partial c}{\partial d}}_{\text{from previous page}} \; \underbrace{\frac{\partial d}{\partial w}}_{x_i}$$

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

$$\underbrace{\frac{\partial d}{\partial w_0}}_{1}$$

Exercise#5:

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

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

$$= \eta \cdot \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot x_i$$

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

STEP #1: Initialize $\underline{w}$, $\underline{w_0}$ and decide ?

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., $\|\Delta w\| \neq 0$, $|\Delta w| \neq 0$]

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