Linear Discrimination

$$\left.\begin{aligned} P(y=1 \mid x) &= \delta \\ P(y=2 \mid x) &= 1-\delta \end{aligned}\right\} \text{ choose } C_1 \text{ if } \begin{cases} \delta > 0.5 \\ \dfrac{\delta}{1-\delta} > 1 \\ \log\left(\dfrac{\delta}{1-\delta}\right) > 0 \end{cases}$$

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

multivariate normal

$N(\mu_1, \Sigma)$ $\quad$ $N(\mu_2, \Sigma)$

shared covariance.

Σ_1 & Σ_2

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

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

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

$$= \Big[\underbrace{\Sigma^{-1}}_{D\times D} . \underbrace{(\mu_1 - \mu_2)}_{D\times 1}\Big]^T . \underbrace{x}_{D\times 1} + \Big[-\frac{1}{2}\underbrace{(\mu_1+\mu_2)^T}_{1\times D} \underbrace{\Sigma^{-1}}_{D\times D} \underbrace{(\mu_1-\mu_2)}_{D\times 1} + \log\Big[\frac{P(y=1)}{P(y=2)}\Big]\Big]$$

$$= \underline{w^T . x} + \underline{w_0}$$

$$w = \widehat{\Sigma^{-1}} . (\widehat{\mu_1} - \widehat{\mu_2})$$

$$w_0 = -\frac{1}{2}(\widehat{\mu_1} + \widehat{\mu_2})^T \widehat{\Sigma^{-1}} . (\widehat{\mu_1} - \widehat{\mu_2}) + \log\widehat{\Big[\frac{P(y=1)}{P(y=2)}\Big]}$$

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

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

$$\underline{\delta = ?}$$

$$\exp\Big[\log\Big(\frac{\delta}{1-\delta}\Big)\Big] = \exp\Big[w^T x + w_0\Big] \Rightarrow \frac{\delta}{1-\delta} = \exp(w^T x + w_0)$$

$$\delta = \exp(w^T x + w_0) - \underbrace{\delta . \exp(w^T x + w_0)}$$

$$\delta(1 + \exp(w^T x + w_0)) = \exp(w^T . x + w_0)$$

$$\boxed{\delta = \frac{\exp(w^T x + w_0)}{1 + \exp(w^T x + w_0)}}$$

(a) ⇒ if $\underline{w^T.x+w_0} > 0 \Rightarrow \delta = \frac{\exp(a)}{1+\exp(a)} > \underline{\underline{0.5}}$

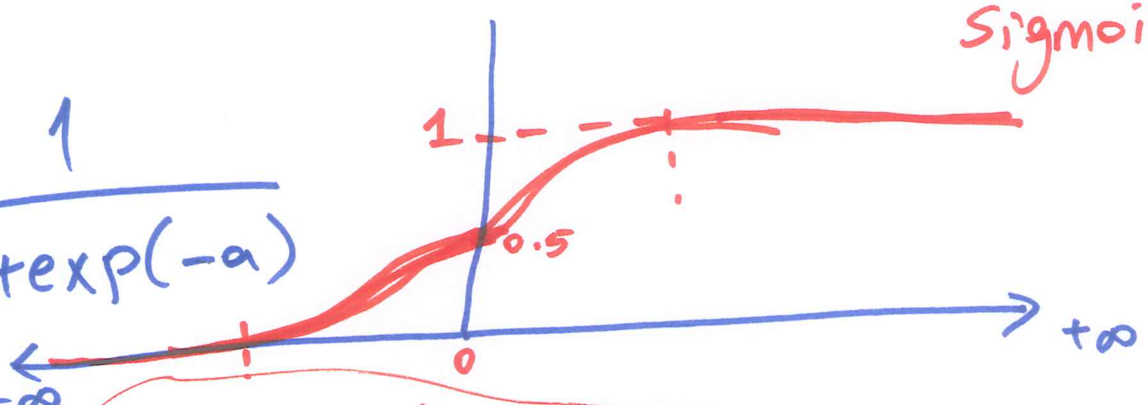
(b) ⇒ if $w^T.x+w_0 = 0 \Rightarrow \underline{\underline{\delta}} = \frac{\exp(0)}{1+\exp(0)} = \frac{1}{2}$

(c) ⇒ if $w^T.x+w_0 < 0 \Rightarrow \delta = \frac{\exp(c)}{1+\exp(c)} < \underline{\underline{0.5}}$

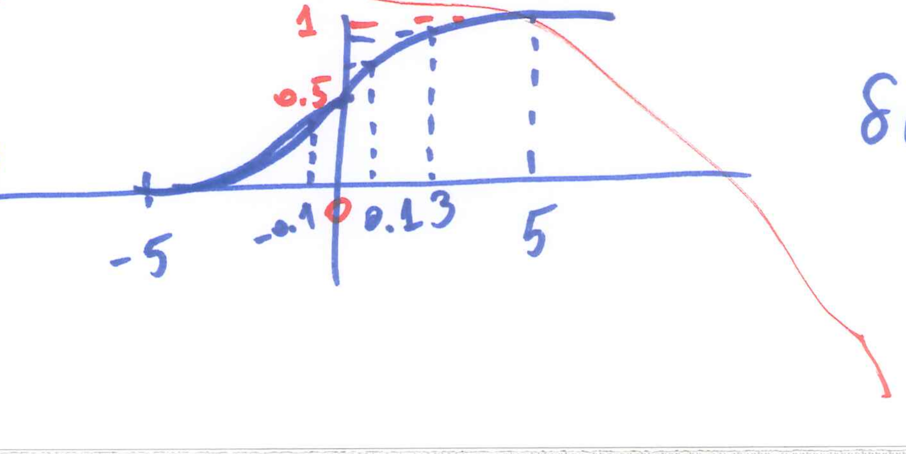
$$\delta = \frac{\exp(w^Tx+w_0)/\exp(w^Tx+w_0)}{[1+\exp(w^Tx+w_0)]/\exp(w^Tx+w_0)} = \boxed{\frac{1}{1+\exp[-(w^Tx+w_0)]}}$$

sigmoid function

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

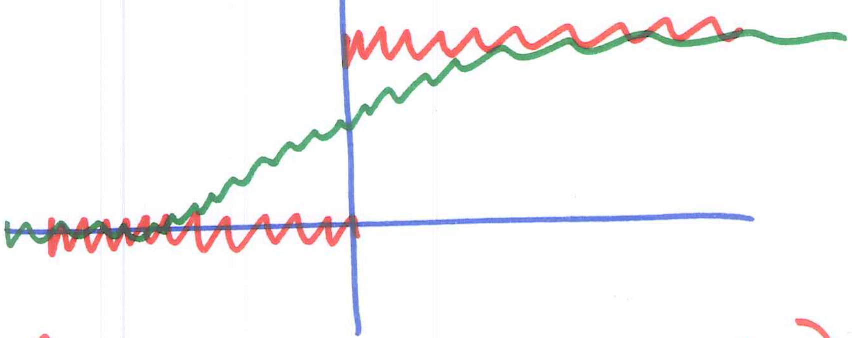


$$\delta(5) = \frac{1}{1+\underbrace{\exp(-5)}} = \frac{1}{1+0.0\ldots\ldots} \simeq 0.999$$

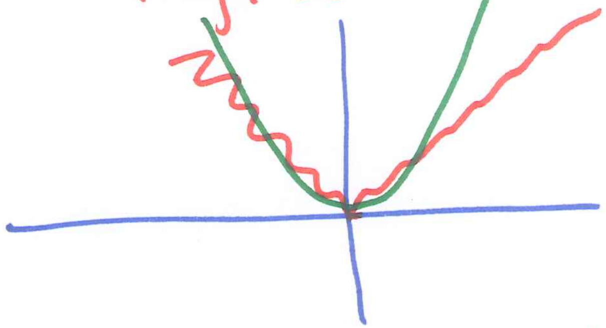


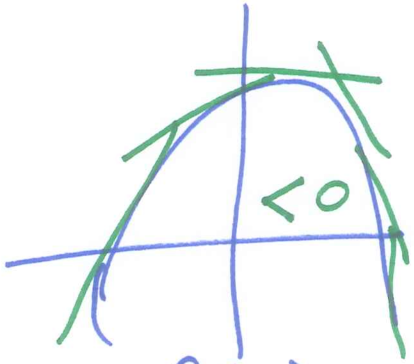
$$\delta(-5) = \frac{1}{1+\underbrace{\exp(5)}_{1000}} = \frac{1}{1+1000} = \frac{1}{1001} \cong 0.00\ldots$$

Classification

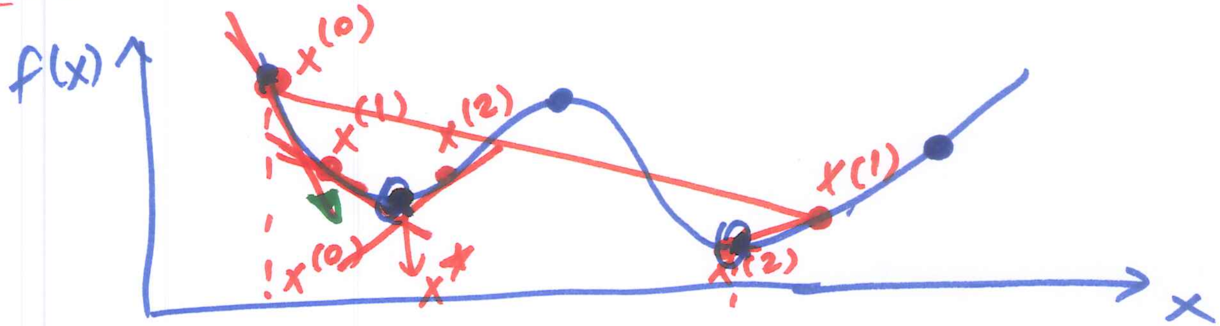


Regression





(Gradient Descent:) (Gradient Ascent)

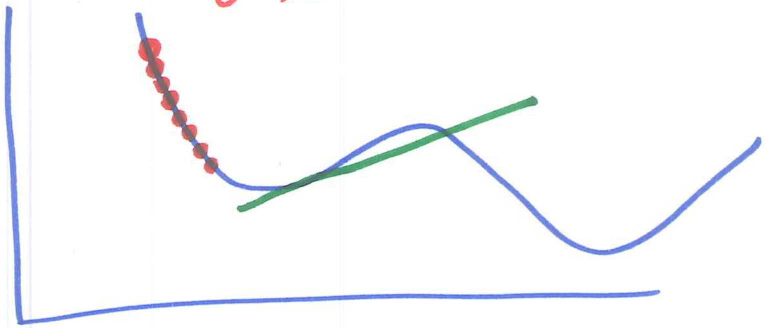


$$x^* = \arg\min_x f(x)$$

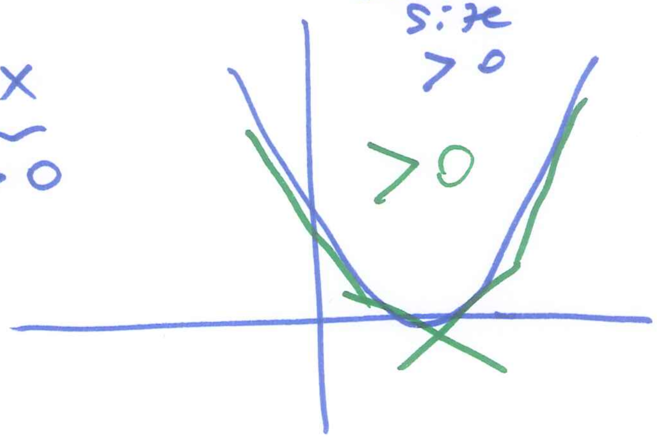
$$\Delta x = -\underbrace{\eta}_{\text{step size} > 0} \underbrace{\frac{\partial f(x)}{\partial x}}_{<0}$$

$$\frac{\partial f(x)}{\partial x}$$

$\Delta x > 0$



$$x^{(1)} = x^{(0)} + \underbrace{\Delta x}_{>0}$$



$$w^* = \arg\min_{w} \underset{\downarrow \atop \text{error}}{E}[w|X] \qquad w = \{w, w_0\}$$

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N} \qquad \underbrace{y_i = 1}_{\text{positive class}} \text{ or } \underbrace{y_i = 0}_{\text{negative class}}$$

probability of success

0 or 1

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

$$y_i | x_i \sim \text{Bernoulli}(\underbrace{\hat{p}(y_i = 1 | x)}))$$

Bernoulli(0.8)

Bernoulli(0.2)

$\hat{p}(y_i = 1 | x) = 0.8$ 0.2

$\hat{p}(y_i = 0 | x) = 0.2$ 0.8

known

$$l(w, w_0 | \mathcal{X}) = \prod_{i=1}^{N} \hat{y}_i^{\,y_i} (1 - \hat{y}_i)^{1 - y_i}$$

estimated

$$\text{Log-Likelihood} = \sum_{i=1}^{N} y_i \log[\hat{y}_i] + (1 - y_i) \log[1 - \hat{y}_i]$$

$$\mathrm{Error}(w, w_0 | \mathcal{X}) = -\,\mathrm{Log\text{-}Likelihood}$$

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

$$\frac{\partial \mathrm{Error}}{\partial w} = ?$$

$$\frac{\partial \mathrm{Error}}{\partial w_0} = ?$$

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

$1 -$

Exercise #4

$$\boxed{\begin{aligned} \mathrm{sigmoid}(a) &= \frac{1}{1+\exp(-a)} \\ \frac{\partial\, \mathrm{sigmoid}(a)}{\partial a} &= \mathrm{sigmoid}(a).[1-\mathrm{sigmoid}(a)] \end{aligned}}$$

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

$$\log[\hat{y}_i] = \log[\underbrace{\text{sigmoid}[\underbrace{w^T \cdot x_i + w_0}_{d}]}_{C}] \qquad w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix}$$

$$\frac{\partial \log[\hat{y}_i]}{\partial w} = \frac{\partial \log[\hat{y}_i]}{\partial c} \frac{\partial c}{\partial d} \frac{\partial d}{\partial w}$$

Exercise#5

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

$$\boxed{\frac{\partial \text{Error}}{\partial w_d} = - \sum_{i=1}^{N} (y_i - \hat{y}_i) \cdot x_{id}}$$

$$\Delta x = -\eta \frac{\partial f(x)}{\partial x}$$

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

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

$$\Delta w_d = -\eta \cdot - \sum_{i=1}^{N} (y_i - \hat{y}_i) x_{id} = \boxed{\eta \sum_{i=1}^{N} (y_i - \hat{y}_i) x_{id}}$$

Step 1 : initialize w & w_0 → initialize them to very small values. uniform $[-0.001, 0.001]$

Step 2 : calculate Δw & Δw_0

Step 3 : update w & w_0 using Δw & Δw_0

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

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

Step 4 : go to $\boxed{\text{step 2}}$ if there is a change in the parameters.

if $|\Delta w| < \epsilon$ & $|\Delta w_0| < \epsilon$

ϵ is a very small number such as 10^{-10}.