Linear Discrimination

$$g(x) = w^T . x + w_0$$

by assuming

$$p(x|y=1) = N(x; \mu_1, \Sigma_1)$$

$$p(x|y=2) = N(x; \mu_2, \Sigma_2)$$

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

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

sample mean vector of the first class

sample mean vector of the second class

$$w_0 = -\frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)\hat{\Sigma}^{-1}(\hat{\mu}_1 - \hat{\mu}_2) + \log\left[\frac{\hat{Pr}(y=1)}{\hat{Pr}(y=2)}\right]$$

sample covariance matrix of all data points

frequency of the first class

frequency of the second class

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

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

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

sigmoid

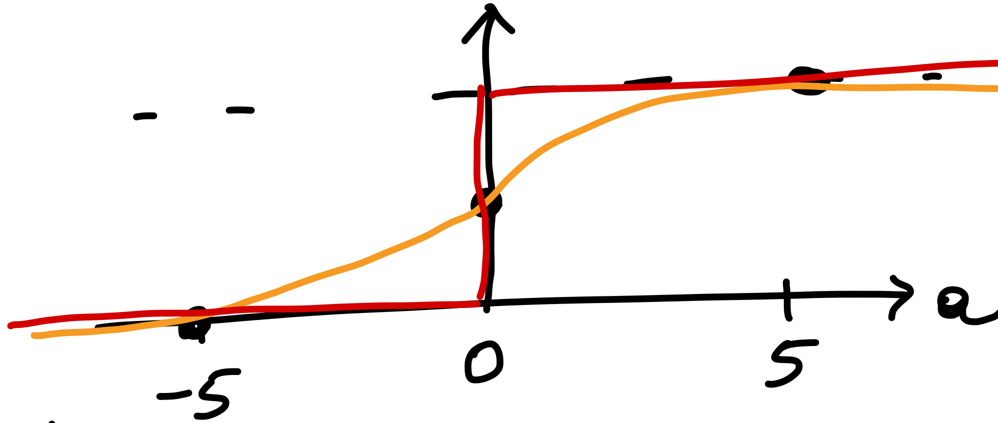
$$\delta = \frac{1}{1 + \exp\left[-(w^T . x + w_0)\right]} \Leftarrow \delta = \frac{\exp\left[w^T . x + w_0\right]}{1 + \exp\left[w^T . x + w_0\right]} \frac{\exp(-w^T . x - w_0)}{\exp(-w^T . x - w_0)}$$

a) if $w^T.x + w_0 > 0 \Rightarrow \delta > 0.5 \Rightarrow 1 - \delta < 0.5$

b) if $w^T.x + w_0 = 0 \Rightarrow \delta = 0.5 \Rightarrow 1 - \delta = 0.5$

c) if $w^T.x + w_0 < 0 \Rightarrow \delta < 0.5 \Rightarrow 1 - \delta > 0.5$

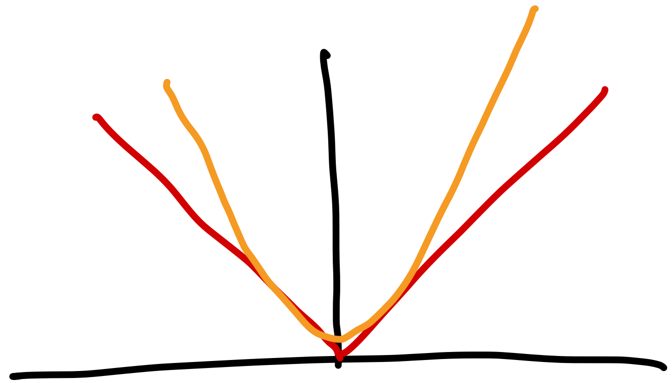
$$\delta(a) = \frac{1}{1+\exp(-a)}$$



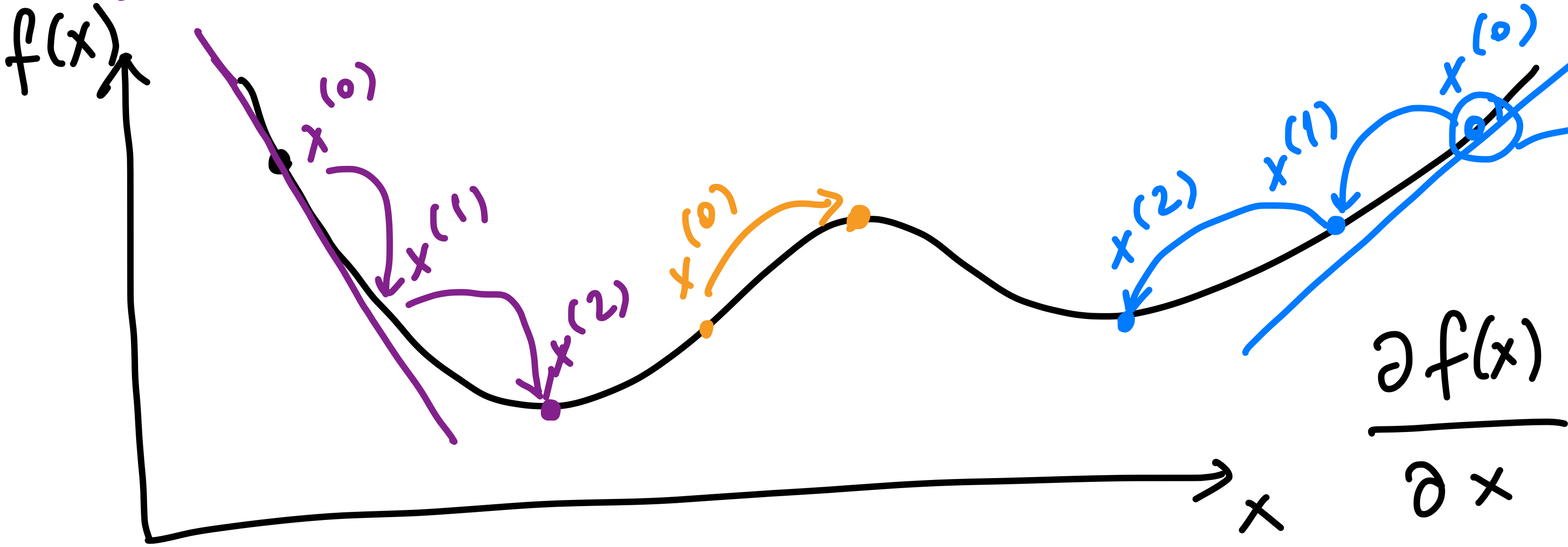
if $a = 5 \Rightarrow \dfrac{1}{1+\exp(-5)} \cong 1$

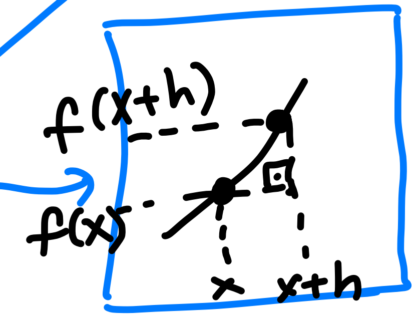
if $a = 0 \Rightarrow \dfrac{1}{1+\exp(0)} = \dfrac{1}{2} = 0.5$

if $a = -5 \Rightarrow \dfrac{1}{1+\exp(+5)} \cong 0$



Gradient Descent / Gradient Ascent





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

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

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

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

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

$\mathcal{X} \to$ training set

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

$1 \to$ success

$0 \to$ failure.

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

$$y_i \mid x_i \sim \text{Bernoulli}(y_i ; \overbrace{\hat{Pr}(y_i = 1 \mid x_i)}^{\hat{y}_i})$$

$\hat{Pr}(y_i = 1 \mid x_i) \to$ predicted posterior prob.

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

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

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

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

$$\text{subject to}: \underbrace{w, w_0}_{\text{unknowns}}$$

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

$$\frac{\partial E[w, w_0 | \mathcal{X}]}{\partial w} = ?$$

$$\frac{\partial E[w, w_0 | \mathcal{X}]}{\partial w_0} = ?$$

$$\frac{\partial \log(\hat{y}_i)}{\partial w} = \frac{1}{\hat{y}_i} \frac{\partial \hat{y}_i}{\partial w}$$

$$\hookrightarrow \frac{\partial \log(\hat{y}_i)}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial w}$$

$$\frac{\partial \log(a)}{\partial a} = \frac{1}{a}$$

$$\frac{\partial \frac{f(a)}{g(a)}}{\partial a} = \frac{\frac{\partial f(a)}{\partial a} \cdot g(a) - \frac{\partial g(a)}{\partial a} \cdot f(a)}{[g(a)]^2}$$

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

$(w^T . x + w_0)$

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

Hint:

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

$$= \frac{\exp(-a)}{[1+\exp(-a)]^2} = \frac{\exp(-a)}{1+\exp(-a)} \frac{1}{1+\exp(-a)}$$

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

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

$$= \frac{1}{\cancel{d}} \quad (\cancel{d})(1-d) \quad x_i$$

$$= (1-\hat{y}_i) \cdot x_i$$

Exercise

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

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

$$\Delta w = -\eta \cdot \frac{\partial Error}{\partial w}$$

$$\Delta w_0 = -\eta \cdot \frac{\partial Error}{\partial w_0}$$