## L1 - Norm Regression model :-

$$\min_{w,b} \frac{\lambda}{2}\|w\|_1 + \frac{1}{L}\sum_{i=1}^{L}\|y_i - (w^T x_i + b)\|_1$$

$$\min_{w,b} \text{Regularisation} + \frac{1}{L}\sum_{i=1}^{L} L(y_i, w, b)$$

23/3/23

# CLASSIFICATION.

If any of the data pts. is very large & all other are very less w.r.t to the largest, then it will effect Euclidean distance & every other operation, so we normalised the data. as,

**Normalisation :-**

For every variable of transaction $x = \dfrac{\text{value of var} - \text{min}^m \text{ value of var.}}{\text{max}^m \text{ value of var.} - \text{min}^m \text{ value of var.}}$

→ * Relative frequency converge to probability for large number of data points.

# Baye's Formula

$$P(w_i/x) = \frac{\overbrace{P(x/w_i)}^{\text{likelihood}}\;\overbrace{P(w_i)}^{\text{Prior prob.}}}{\underbrace{P(x)}_{\text{evidence}}}$$

if $P(w_1/x) \geq P(w_2/x)$ ; decide $w_2$

other, decide $w_2$.

if $P(x/w_1)P(w_1) \geq P(x/w_2)P(w_2)$ ; decide $w_1$

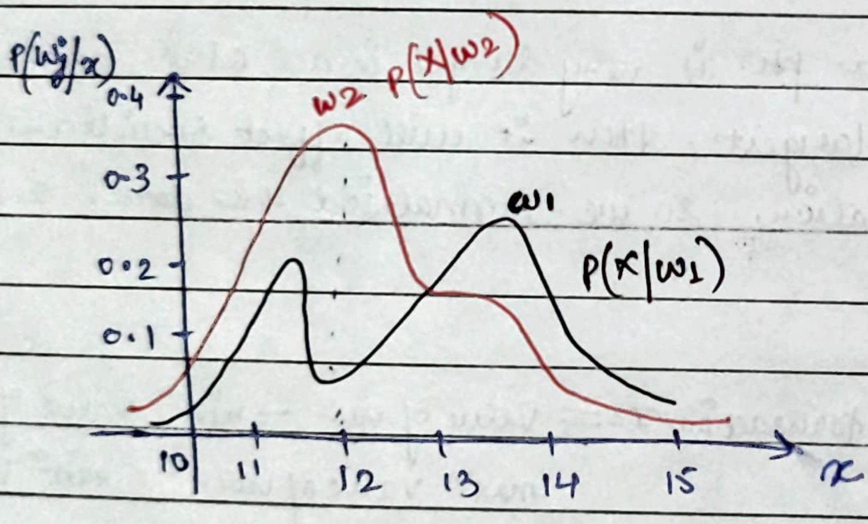
other, decide $w_2$

$P(x) = P(x/w_1)\,P(w_1) + P(x/w_2)\,P(w_2) \longrightarrow$ Total prob.

→ Bayes formula can be written as:-

$$\boxed{p(x) = \sum_{j=1}^{2} p(x/w_j)\,P(w_j)}$$

$$\boxed{\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}}$$




at $x=12$, density of $w_2$ is high, so we will consider $w_2$ at $x=12$.

* if $P(w_j/x) \geq P(w_i/x)$ ; decide $j^{th}$ class
where $i \neq j$ ; $i = 1:K$

→ $\boxed{\text{Error} = 1 - P(w_j/x)}$

Duda & Hearts Book
ch-2 → pattern classification pdf

* $P(error/x) = \begin{cases} P(w_1/x) & ; \text{ if we decide } w_2 \\ P(w_2/x) & ; \text{ if we decide } w_1 \end{cases}$

* $P(error) = \int_{-\infty}^{\infty} P(error, x)\, dx = \int_{-\infty}^{\infty} P(error/x)\, p(x)\, dx.$

**Baye's Decision Rule:**

* Decide $w_1$, if $P(w_1/x) > P(w_2/x)$

otw Decide $w_2$

* $P(error/x) = \min[P(w_1/x), P(w_2/x)]$

3/4/23

$$P(w_j/x) = \frac{P(x/w_j)\,P(w_j)}{P(x)}$$

Decide

$w_j$ class if $P(w_j/x) \geq P(w_i/x)$ $\forall\, i = 1, 2, \dots c$ and $i \neq j$

$\Rightarrow P(x/w_j)\,P(w_j) \geq P(x/w_i)$ for $i = 1:c$, $i \neq j$

$g_i(x)$ → tells prob. of ~~$i^{th}$ place to to belong~~ ~~in which class~~.

$\boxed{\text{Discriminant fn.}}$

classify, $g_j(x) \geq g_i(x)$ $\forall\, i = 1, 2, \dots C$, $i \neq j$

$$g_i(x) = \log(P(x/w_i)\ P(w_i))$$

$$= \log P(x/w_i) + \log P(w_i)$$

Let $c = 2$ (Binary class)

then, class 1 $g_1(x) = \log P(x/w_1) + \log P(w_1)$

& class 2 $g_2(x) = \log P(x/w_2) + \log P(w_2)$.

calculate $g(x) = g_1(x) - g_2(x)$

Decide class 1 if $g(x) \geq 0$.
otw class 2.

$$g(x) = \log \frac{P(x/w_1)}{P(x/w_2)} + \log \frac{P(w_1)}{P(w_2)}$$

if $X \in \mathbb{R}^n$ then $\mu \in \mathbb{R}^n$, & $\hat{\mu} = \bar{x}$

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{r1} & \sigma_{r2} & \cdots & \sigma_{rn} \end{bmatrix}$$

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\frac{1}{\sigma 2\sqrt{\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{2}}$$

Date ______
Page ______

$\Sigma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ → var. is same in dist for $x_1$ & $x_2$

$Cov(x_1, x_2) = 0$

Hence, density graph is symmetric and countour plot is concentric circles.

- Case I: $\Sigma_i = \sigma^2 I = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ → we get linear classifier

$$g_i(x) = \log P(x/w_i) + \log P(w_i)$$

$$= \log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i)\right]\right) + \log P(w_i)$$

d = dimension

$|\Sigma| = \sigma^2 \Rightarrow |\Sigma|^{1/2} = \sigma$

$$= -\frac{1}{2}(x-\mu_i)^T \Sigma^{-1}(x-\mu_i) + \log P(w_1)$$ ← ∵ $\log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right)$ is same in both $x_1$ & $x_2$, so we take rest terms

$\Sigma^{1/2} = \frac{1}{\sigma^2}$

$$= \frac{-(x-\mu_i)^T(x-\mu_i)}{2\sigma^2} + \log P(w_i)$$

$\|y\|^2 = y^T y$

$$g_i(x) = -\frac{\|x-\mu_i\|^2}{2\sigma^2} + \log P(w_i)$$

$$= \frac{-1}{2\sigma^2}\left[x^T x - 2\mu_i^T x + \mu_i^T \mu_i\right] + \log P(w_i)$$

$$g_i(x) = \frac{-1}{2\sigma^2}\left[-2\mu_i^T x + \mu_i^T \mu_i\right] + \log P(w_i)$$

$$= w_i^T x + w_{i0}$$

$w_i = \frac{1}{\sigma^2} u_i$

$w_0 = \frac{-1}{2\sigma^2} u_i^T u_i + \log P(w_i)$

$\because$ we have 2 class, then,

$g(x) = g_1(x) - g_2(x)$

if $g(x) \geq 0$, then class 1.

$$g_1(x) - g_2(x) = \frac{-1}{2\sigma^2}\left[-2\mu_1^T x + \mu_1^T \mu_1\right] + \log P(\omega_1)$$

$$+ \frac{1}{2\sigma^2}\left[-2\mu_2^T x + \mu_2^T \mu_2\right] - \log P(\omega_2)$$

$$= \frac{\mu_1^T x}{\sigma^2} - \frac{\mu_1^T \mu_1}{2\sigma^2} + \log P(\omega_1) - \frac{\mu_2^T x}{\sigma^2} + \frac{\mu_2^T \mu_2}{2\sigma^2} - \log P(\omega_2)$$

$$= \frac{x}{\sigma^2}(\mu_1^T - \mu_2^T) + \underbrace{\frac{1}{2\sigma^2}(\mu_2^T \mu_2 - \mu_1^T \mu_1) + \log \frac{P(\omega_1)}{P(\omega_2)}}_{w_{10}}$$

$$= \frac{1}{\sigma^2}(\mu_1^T - \mu_2^T) x + w_{10}$$

$$= w^T x + w_{10}.$$

5/4/23

$$g(x) = \frac{1}{\sigma^2}(\mu_1 - \mu_2)^T x + \frac{1}{2\sigma^2}(\mu_2^T \mu_2 - \mu_1^T \mu_1) + \log \frac{P(\omega_1)}{P(\omega_2)} = 0$$

$$\Rightarrow (\mu_1 - \mu_2)^T x + \frac{1}{2}(\mu_2^T \mu_2 - \mu_1^T \mu_1) + \log \frac{P(\omega_1)}{P(\omega_2)} \sigma^2 = 0$$

$$\Rightarrow w^T x + w_{10} = 0$$

$$w = \mu_1 - \mu_2$$

$$\Rightarrow (\mu_1 - \mu_2)^T x + \frac{1}{2}(\mu_2 - \mu_2)^T(\mu_1 + \mu_2) + \frac{\sigma^2}{\|\mu_1 - \mu_2\|^2}(\mu_1 - \mu_2)^T(\mu_1 - \mu_2) \log \frac{P(\omega_1)}{P(\omega_2)}$$

$$\Rightarrow (\mu_1-\mu_2)^T\left[x+\frac{1}{2}(\mu_2 \mp \mu_1) + \frac{\sigma^2}{\|\mu_1-\mu_2\|} \log\frac{P(w_1)}{P(w_2)}(\mu_1-\mu_2)\right.$$

$$\Rightarrow w^T\left[x-\frac{1}{2}(\mu_1+\mu_2)+\frac{\sigma^2}{\|\mu_1-\mu_2\|^2}\log\frac{P(w_1)}{P(w_2)}(\mu_1-\mu_2)\right.$$

$$\rightarrow w_0 = \mu_1-\mu_2$$

$$x_0 = \frac{1}{2}(\mu_1+\mu_2) - \frac{\sigma^2}{\|\mu_1-\mu_2\|^2}\log\frac{P(w_1)}{P(w_2)}(\mu_1-\mu_2)$$

$$\rightarrow g(x) = w^T(x-x_0) = 0$$

we get linear classifier

- Case II : $\Sigma_1 = \Sigma_2 = [\quad]$ → all entries are non-zero.

$$g_i(x) = \log P(X/w_1) + \log P(w_1)$$

$$= \log\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\right) + \log\left(e^{-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)}\right)$$

$$+ \log P(w_1)$$

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i) + \log P(w_1)$$

$$\rightarrow g_1(x) = -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \log P(w_1)$$

$$\rightarrow g_2(x) = -\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) + \log P(w_1)$$

$$g(x) = g_1(x) - g_2(x)$$

$$\Rightarrow -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \log P(w_1) + \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) - \log P(w_1)$$

$$\Rightarrow -\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2) + \log\frac{P(w_1)}{P(w_2)}$$

$$= \frac{1}{2}\left[ (x-\mu_2)^T \Sigma^{-1} (x-\mu_2) - (x-\mu_1)^T \Sigma^{-1} (x-\mu_1) \right] + \log \frac{P(\omega_1)}{P(\omega_2)}$$

$$\cancel{\Rightarrow \frac{1}{2}\Sigma^{-1}\left[ (x-\mu_2)^T (x-\mu_2) - (x-\mu_1)^T (x-\mu_1) \right] + \log \frac{P(\omega_1)}{P(\omega_2)}}$$

$$\cancel{\Rightarrow \frac{1}{2}\Sigma^{-1}\left[ (x-\mu_2)^2 - (x-\mu_1)^2 \right] + \log \frac{P(\omega_1)}{P(\omega_2)}}$$

$$\Rightarrow \frac{1}{2}\Big[ \cancel{x^T \Sigma^{-1} x} - x^T \Sigma^{-1} \mu_2 - \mu_2^T \Sigma^{-1} x + \mu_2^T \Sigma^{-1} \mu_2$$

$$- \left( \cancel{x^T \Sigma^{-1} x} - x^T \Sigma^{-1} \mu_1 - \mu_1^T \Sigma^{-1} x + \mu_1^T \Sigma^{-1} \mu_1 \right) \Big] + \log \frac{P(\omega_1)}{P(\omega_2)}$$

$$\Rightarrow \frac{1}{2}\Big[ - x^T \Sigma^{-1} \mu_2 - \mu_2^T \Sigma^{-1} x + \mu_2^T \Sigma^{-1} \mu_2 + x^T \Sigma^{-1} \mu_1 + \mu_1^T \Sigma^{-1} x$$

$$- \mu_1^T \Sigma^{-1} \mu_1 \Big] + \log \frac{P(\omega_1)}{P(\omega_2)}$$

$\omega^T(x - x_0)$

$$\cancel{\Rightarrow \frac{1}{2}\Big[ (\mu_1 - \mu_2) \; \mu_1 \left( x^T \Sigma^{-1} + \Sigma^{-1} x \right) - \mu_1^T \Sigma^{-1} x \right)}$$

$$\cancel{- \mu_2 \left( x^T \Sigma^{-1} + \mu_2^T \Sigma^{-1} + \mu \right.}$$

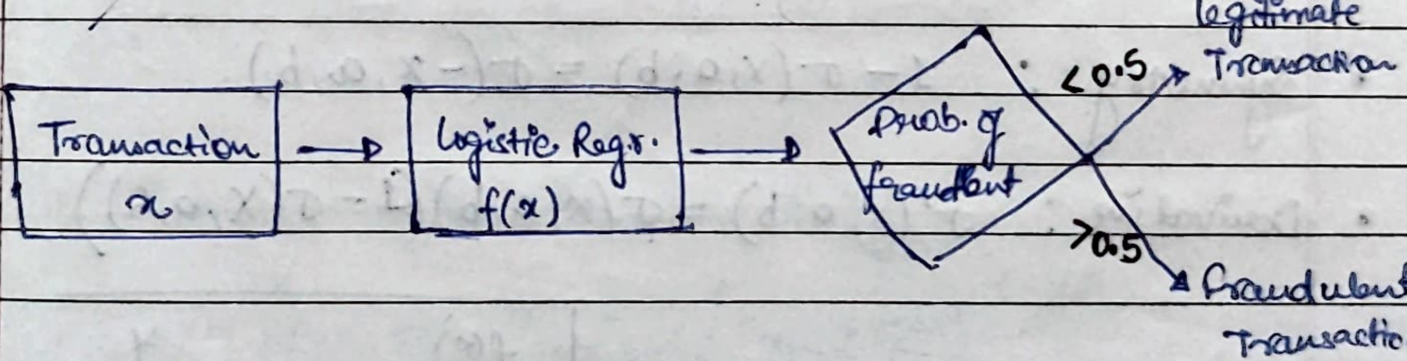
- case III : $\Sigma_1$ and $\Sigma_2$ are not equal → we get here non-linear classifier

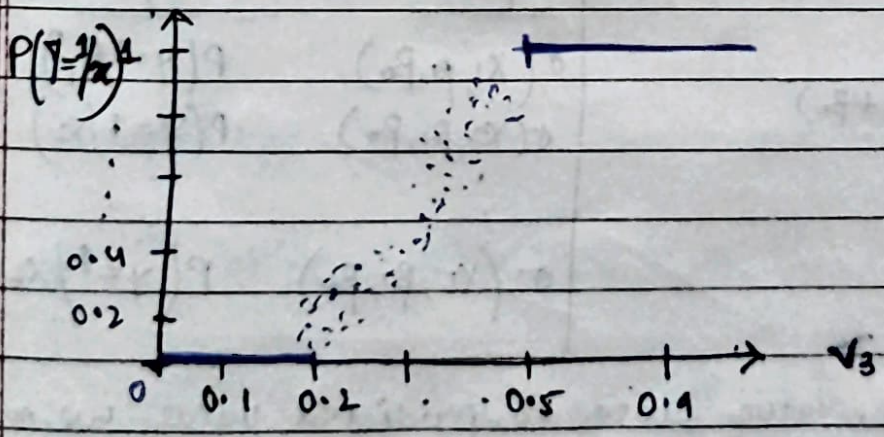
$$g_1(X) = -\frac{1}{2}(X-\mu_1)^T \Sigma_1^{-1}(X-\mu_1) + \log P(w_1) - \frac{1}{2}\log|\Sigma_1|$$

$$g_2(X) = -\frac{1}{2}(X-\mu_2)^T \Sigma_2^{-1}(X-\mu_2) + \log P(w_2) - \frac{1}{2}\log|\Sigma_2|$$

# Logistic Regression

- Non-parametric Regression Model is Logistic Regression.




from (0 – 0.2) prob. of fraudlent is 0.
& from (0.5 – 1) " " " " 1.
in b/w, we have to fit a curve.

# Sigmoidal function.

$$\frac{1}{1+e^{-(\beta_1 x+\beta_0)}} = \frac{e^{(\beta_1 x+\beta_0)}}{1+e^{(\beta_1 x+\beta_0)}}$$

↑ value of $\beta_1$ → we get change in Y. (in graph)
↑ value of $\beta_0$ → we shift the fn..
ie $\beta_0$ is shift operator.

- $\sigma(x, a, b) = \frac{1}{1+e^{-(\beta_1 x+\beta_0)}} = \frac{e^{\beta_1 x+\beta_0}}{1+e^{\beta_1 x+\beta_0}}$

- Symmetry: $1-\sigma(x,a,b) = \sigma(-x,a,b)$.

- Derivative: $\sigma'(x,a,b) = \sigma(x,a,b)(1-\sigma(x,a,b))$.

$$\sigma(x,a,b) = \frac{1}{1+e^{-(\beta_1 x+\beta_0)}}$$

| f(x) | Y |
|---|---|
| $\sigma(x_1, \beta_1, \beta_0)$ | $P(y=1/x_1)$ |
| $\sigma(x_2, \beta_1, \beta_0)$ | $P(y=1/x_2)$ |
| $\vdots$ | $\vdots$ |
| $\sigma(x_L, \beta_1, \beta_0)$ | $P(y=1/x_L)$ |

To get actual value close to predicted value, we minimise the loss fn using least square technique.

ie. $$\min_{(\beta_1,\beta_0)} \sum_{i=1}^{L} \left(Y_i - \sigma(x_i, \beta_1, \beta_0)\right)^2$$

Date
Page

$T = \{(x_1, y_1), (x_2, y_2), \dots\dots, (x_L, y_L)\}$

$y_i = 1$

$y_i = 0$

$(x_1, 1), (x_2, 0), (x_3, 1), (x_4, 1), (x_5, 1), (x_6, 0), \dots$

$\prod_{i: y_i=1} P(x_i) \cdot \prod_{i: y_i=0} 1 - P(x_i)$

$P(x_i) = P(y_i = 1/x_i)$
$P(y_i = 0/x_i) = 1 - P(y_i = 1/x_i)$

or

Maximize $\prod_{i: y_i=1} P(y_i = 1/x_i) \cdot \prod_{i: y_i=0} 1 - P(y_i = 1/x_i)$

ie, we have to maximise the probability.

$$\text{Max} \prod_{i=1}^{n} P(y_i = 1/x_i)^{y_i} \cdot (1 - P(y_i = 1/x_i))^{1-y_i}$$

Taking log,

$$= \log \prod_{i=1}^{n} P(y_i = 1/x_i)^{y_i} \cdot (1 - P(y_i = 1/x_i))^{1-y_i}$$

$$= \sum_{i=1}^{n} \left[ \log P(y_i = 1/x_i) \cdot y_i + (1-y_i) \log (1 - P(y_i = 1/x_i)) \right]$$

Maximize cross Entropy loss function

$$\text{Max} \left[ \sum_{i=1}^{n} y_i \log \underbrace{P(y_i = 1/x_i)}_{\sigma(x_i, \beta_1, \beta_0)} + (1-y_i) \log (1 - P(y_i = 1/x_i)) \right]$$

$$\max_{\beta_1, \beta_0} \left[ \sum_{i=1}^{n} y_i \log \sigma(x_i, \beta_1, \beta_0) + (1-y_i) \log (1 - \sigma(x_i, \beta_1, \beta_0)) \right]$$

we have to minimise this loss fn.

$$y_i(\beta_1 x+\beta_0) - \log(1+e^{\beta_1 x+\beta_0})$$

Date ______
Page ______

$$\max \sum y_i \log\left(\frac{e^{\beta_1 x+\beta_0}}{1+e^{\beta_1 x+\beta_0}}\right) + (1-y_i)\log\left(1-\frac{e^{\beta_1 x+\beta_0}}{1+e^{+\beta_1 x+\beta_0}}\right)$$

$$\max \sum y_i\left[(\beta_1 x+\beta_0) - \log(1+e^{\beta_1 x+\beta_0}) + (1-y_i)\log\left(\frac{1+e^{-\beta_1 x+\beta_0}-e^{\beta_1 x+\beta_0}}{1+e^{+\beta_1 x+\beta_0}}\right)\right]$$

$$\max \sum y_i(\beta_1 x+\beta_0) - \log(1+e^{\beta_1 x+\beta_0}) + (1-y_i)(-\log(1+e^{\beta_1 x+\beta_0}))$$

$$\Rightarrow \cancel{\sum y_i \beta_1 x \log(1+e^{\beta_1 x+\beta_0}) + y_i\beta_0 \log}$$

$$= \cancel{\sum_{i=1}^{n} \log(1+e^{\beta_1 x+\beta_0})\left[y_i(\beta_1 x+\beta_0) - (1-y_i)\right]}$$

$$\cancel{\max \sum y_i(\beta_1 x+\beta_0) - \log(1+e^{\beta_1 x+\beta_0}) - \log(1+e^{\beta_1 x+\beta_0}) + y_i\log(1+e^{\beta_1 x+\beta_0})}$$

$$\Rightarrow \max \sum \left(y_i \log(e^{\beta_1 x+\beta_0}) - y_i\log(1+e^{\beta_1 x+\beta_0}) + (1-y_i)\log\left(\frac{1+e^{\beta_1 x+\beta_0}-e^{\beta_1 x+\beta_0}}{1+e^{\beta_1 x+\beta_0}}\right)\right.$$

$$\Rightarrow \max \sum y_i(\beta_1 x+\beta_0) - y_i\log(1+e^{\beta_1 x+\beta_0}) + (1-y_i)(-\log(1+e^{\beta_1 x+\beta_0}))$$

$$\sum y_i(\beta_1 x+\beta_0) - \cancel{y_i\log(1+e^{\beta_1 x+\beta_0})} - \log(1+e^{\beta_1 x+\beta_0}) + \cancel{y_i\log(1+e^{\beta_1 x+\beta_0})}$$

$$\Rightarrow \boxed{\max \sum y_i(\beta_1 x+\beta_0) - \log(1+e^{\beta_1 x+\beta_0})}$$

Date
Page

$$L(y, f(x)) = \begin{cases} -\log(f(x)) , & \text{if } y = 1 \\ -\log(1-f(x)) , & \text{if } y = 0 \end{cases}$$

$$= -y\log(f(x)) - (1-y)\log(1-f(x))$$

$$\min_{f} \sum_{i=1}^{L} -y_i \log(f(x_i)) - (1-y_i)\log(1-f(x_i))$$

$$\min_{\beta_1,\beta_0} \sum_{i=1}^{L} -y_i \log\left(\frac{1}{1+e^{-(\beta_1 x+\beta_0)}}\right) - (1-y_i)\log\left(1 - \frac{1}{1+e^{-(\beta_1 x+\beta_0)}}\right)$$

$$\left[ f = \min_{\beta_1,\beta_0} -\sum_{i=1}^{L}\left( y_i(\beta_1 x_i+\beta_0) - \log\left(1+e^{(\beta_1 x_i+\beta_0)}\right)\right). \right]$$

$$-\frac{\nabla\partial f}{\partial\beta_1} = y_i(\cancel{\beta_1}\, x_i) + (\beta_1 x_i+\beta_0)(0) - \frac{1}{1+e^{(\beta_1 x_i+\beta_0)}}\beta_1 (e^{\beta_1 x_i+\beta_0})$$

$$= y_i \cancel{\beta_1}\, x_i - \frac{\cancel{\beta_1}\, e^{\beta_1 x_i+\beta_0}}{1+e^{\beta_1 x_i+\beta_0}} \cdot x_i$$

in matrix form

wrt $\beta_1$
$$\boxed{-\frac{\nabla\partial f}{\partial\beta_1} = -\sum_{i=1}^{n}\left(y_i - \frac{e^{\beta_1 x_i+\beta_0}}{1+e^{\beta_1 x+\beta_0}}\right)x_i} = X^T(u-Y)$$

$$= y_i - \sigma(X,\beta_1,\beta_0)x_i$$

$e = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

eigenvector of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$-\frac{\nabla\partial f}{\partial\beta_0} = y_i - \frac{1}{1+e^{\beta_1 x_i+\beta_0}} e^{(\beta_1 x_i+\beta_0)}$$

$= e^T(u-Y)$

in matrix form

wrt $\beta_0$
$$\boxed{+\frac{\nabla\partial f}{\partial\beta_0} \sum_{i=1}^{n}\left(y_i - \frac{e^{\beta_1 x_i+\beta_0}}{1+e^{\beta_1 x_i+\beta_0}}\right) \Rightarrow -\sum_{i=1}^{n}\left(y_i - \sigma(X,\beta_1,\beta_0)\right)}$$

$\beta_1 x_i \rightarrow \underset{(\text{matrix})}{\beta_1^T x_i + \beta_0}$

where, $u = \begin{bmatrix} \sigma(x_1, \beta_1, \beta_0) \\ \sigma(x_2, \beta_1, \beta_0) \\ \vdots \\ \sigma(x_L, \beta_1, \beta_0) \end{bmatrix}$

<u>6/4/23</u> $P(Y=1/x) = \dfrac{1}{1+e^{-\beta_1 x+\beta_0}} = \dfrac{e^{\beta_1 x+\beta_0}}{1+e^{\beta_1 x+\beta_0}}$

$$\sum_{i=1}^{n}\left(y_i - \frac{e^{\beta_1 x+\beta_0}}{1+e^{\beta_1 x+\beta_0}}\right)^2$$

# <u>Minimising cross Entropy loss $f^n$</u>

$$J(\beta_1,\beta_0) = \min_{\beta_1,\beta_0} - \sum_{i=1}^{n}\left(y_i(\beta_1 x_i+\beta_0) - \log(1+e^{(\beta_1 x_i+\beta_0)}\right)$$

# <u>Taking Gradient wrt $\beta_1$ & $\beta_0$</u>

$$\nabla_{\beta_1} J(\beta_1,\beta_0) = - \sum_{i=1}^{n}\left(y_i - \frac{e^{\beta_1 x_i+\beta_0}}{1+e^{\beta_1 x_i+\beta_0}}\right) \cdot x_i$$

$$\nabla_{\beta_0} J(\beta_1,\beta_0) = - \sum_{i=1}^{n}\left(y_i - \frac{e^{\beta_1 x_i+\beta_0}}{1+e^{\beta_1 x_i+\beta_0}}\right)$$

* $P(y_i=1/x_i) = \dfrac{e^{\beta_1^T x+\beta_0}}{1+e^{\beta_1^T x+\beta_0}} = \dfrac{1}{1+e^{-\beta_1^T x+\beta_0}}$

* $P(y_i=0/x_i) = 1 - P(y_i=1/x_i)$

$$= 1 - \frac{1}{1+e^{-\beta_1^T x+\beta_0}} = \frac{e^{-\beta_1^T x+\beta_0}}{1+e^{-\beta_1^T x+\beta_0}}$$

$$= \frac{1}{1+e^{\beta_1^T x+\beta_0}}$$

$$\frac{P(Y=1/x_i)}{1-P(Y=1/x_i)} = e^{\beta^T x + \beta_0}$$

$$\log\left(\frac{P(Y=1/x_i)}{1-P(Y=1/x_i)}\right) = \beta^T x + \beta_0$$

$$\Rightarrow \log\left(\frac{P(Y=1/x_i)}{P(Y=0/x_i)}\right) = \beta^T x + \beta_0$$

$$\Rightarrow \quad g(x) = \begin{cases} 0 \; ; \text{ if } \dfrac{e^{\beta^T x + \beta_0}}{1 + e^{\beta^T x + \beta_0}} \leq 0.5 \\ 1 \; ; \text{ if otw} > 0.5 \end{cases}$$

$$\frac{e^{\beta^T x + \beta_0}}{1 + e^{\beta^T x + \beta_0}} = 0.5$$ (iff $\beta^T x + \beta_0 = 0$, then only we get $0.5$ $(=1/2)$)

$$\Rightarrow \beta^T x + \beta_0 = 0.$$

ie,
$$g(x) = \begin{cases} 0 \; ; \text{ if } \beta^T x + \beta_0 < 0 \\ 1 \; ; \text{ if } \beta^T x + \beta_0 > 0 \end{cases}$$

* Linear fn. ie $f(x)$, here does not fit the data properly. So, we need more complex fn. ie, we take quadratic fn. ie, Non-Linear Logistic function.

# Non-Linear Logistic Regression

$$f(x) = \frac{1}{1+e^{-(\beta_1 x_1^2 + \beta_2 x_2^2 + \beta_3 x_1 x_2 + \beta_4 x_1 + \beta_5 x_2 + \beta_6)}}$$

$$\frac{e^{\sigma(\beta^T x + \beta_0)}}{1+e^{\sigma(\beta^T x + \beta_0)}}$$

**Example:** Consider NAND dataset. Which one of the following may be the sol^n of logistic regression model?

| | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| $x_1$ | 0 | 0 | 1 |
| $x_2$ | 0 | 1 | 1 |
| $x_3$ | 1 | 0 | 1 |
| $x_4$ | 1 | 1 | 0 |

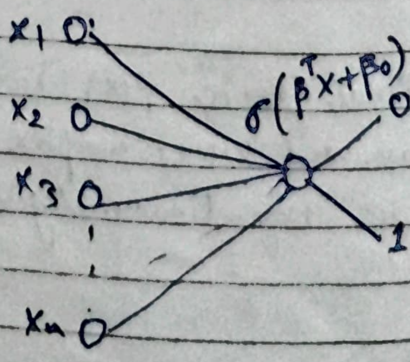
(a) $\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} -80.26 \\ -110.85 \\ 21 \end{bmatrix}$

(b) $\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} 80.26 \\ 110.85 \\ -152.61 \end{bmatrix}$

(c) $\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} -80.26 \\ 0 \\ 152.61 \end{bmatrix}$

(d) $\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} -80.26 \\ -110.85 \\ 152.61 \end{bmatrix}$

$$\sigma(\beta^T x + \beta_0) = \frac{1}{1+e^{-(\beta_1^T x + \beta_0)}}$$



SVM

PCA

KMeans Clustering

Date ____
Page ____

for $X_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ , $\beta^T = [\beta_1 \quad \beta_2]$

(a) $\sigma(\beta^T x + \beta_0) = \dfrac{1}{1+e^{-(\beta_1^T x + \beta_2^T x + \beta_0)}} = \dfrac{1}{1+e^{-21}} = 0.99$

(b) $\dfrac{1}{1+e^{-\beta_0}} = \dfrac{1}{1+e^{152.61}}$

(c) $= \dfrac{1}{1+e^{-152.61}} = 1.$

(d) $\dfrac{1}{1+e^{-152.61}} = 1$

for $X_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ , $\sigma(\beta^T x + \beta_0) = \dfrac{1}{1+e^{-(0 + \beta_2^T x + \beta_0)}}$

(a) $\dfrac{1}{1+e^{-(-110.85 + 21)}} = 9.52 \times 10^{-40}$ $(< 0.5)$

(b) $\dfrac{1}{1+e^{-(110.85 - 152.61)}} = 7.30 \times 10^{-19}$

(c) $\dfrac{1}{1+e^{-(0+152.61)}} = 1$

(d) $\dfrac{1}{1+e^{-(-110.85 + 152.61)}} = 1.$

⋮