Logistic Regression

$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$

y_i = label / Class = $\{0, 1, \ldots, K-1\}$

M/W, P/N

$K = 2$

$\{0, 1\}$

$f_\theta(x) = y$

$P_\theta(y = 1 \mid x) = .9$

$P_\theta(y = 0 \mid x) = 1 - P(y = 1 \mid x) = .1$

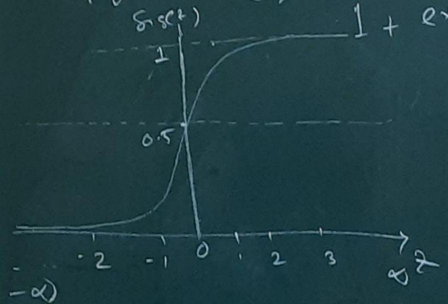
$f_\theta(x) = \theta^T x$ – Unconstrained real number

$-\infty$ ∞

How do we bound $f_\theta(x)$ s.t. it produces values betw. 0 to 1

Logistic/Sigmoid/Expit $\quad \mathcal{D} = \{(x_i, y_i)\} \quad y_i \in \{0, 1\}$

$$\text{Sigmoid}(z) = \frac{1}{1 + \exp(-z)} = \frac{\exp(z)}{1 + \exp(z)}$$



$$f_\theta(x) = \theta^T x \quad \text{– Unconstrained real number}$$

$$-\infty \quad \infty$$

How do we bound $f_\theta(x)$ s.t. it produces values betn. 0 to 1

$$\boxed{p_\theta(y=1|x) = \text{Sigmoid}(f_\theta(x))}$$

MLE of θ

$$\arg\max \prod_{i=1}^{N} \underbrace{p_\theta(1|x_i)}_{y_i=1} \; \underbrace{p_\theta(0|x_i)}_{y_i=0}$$

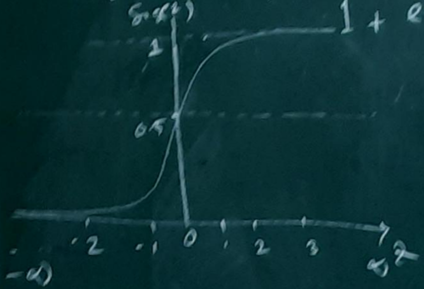
$$= \arg\max \prod_{i=1}^{N} p_\theta(0|x_i)^{1-y_i} \; p_\theta(1|x_i)^{y_i}$$

σ (Sig) = Sig (1 - Sig)

Logistic/Sigmoid/Expit

$\mathcal{D} = \{(x_i, y_i)\}$ $\quad y_i \in \{0,1\}$

$$\text{Sigmoid}(z) = \frac{1}{1+\exp(-z)} = \frac{\exp(z)}{1+\exp(z)}$$



$f_\theta(x) = \theta^T x$ - Unconstrained real numb.

$-\infty \quad \infty$

How do we bound $f_\theta(x)$ s.t it produce values betn 0 to 1

$$\boxed{p_\theta(y=1|x) = \text{Sigmoid}(f_\theta(x))}$$

MLE of θ

$$\arg\max \prod_{i=1}^{N} \underbrace{p_\theta(1|x_i)}_{y_i=1} \; \underbrace{p_\theta(0|x_i)}_{y_i=0}$$

$$= \arg\max \prod_{i=1}^{N} p_\theta(0|x_i)^{1-y_i} \; p_\theta(1|x_i)^{y_i}$$

$$= \arg\max \sum_{i=1}^{N} (1-y_i) \underbrace{\log p_\theta(y=0|x_i)}_{-\log\{\exp \theta^T x_i + 1\}} + y_i \underbrace{\log p_\theta(y=1|x_i)}_{\theta^T x_i - \log\{\exp \theta^T x_i + 1\}}$$

Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$$\arg\max_{\theta} \sum_{i=1}^{N} (1-y_i)\{-\log\{\exp(\theta^T x_i)+1\} + y_i(\theta^T x_i - \log\{\exp\theta^T x_i + 1\})$$

probabilistic

Class conditionals

Discriminative

$p_\theta(y|x)$

forget about $p(x)$

Generative

Class prior

$p_\theta(x|y)$ $p_\theta(y)$

$$p_\theta(y|x) = \frac{p_\theta(x|y)\, p(y)}{p(x)}$$

$+1\}$

Sigmoid / Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$(1-y_i)\{-\log\{\exp(\theta^T x_i)+1\}\} + y_i(\theta^T x_i - \log\{\exp \theta^T x_i + 1\})$

LDA : Linear Discriminant Analysis

assump.

$p_\theta(y=0) = p_\theta(y=1) = 0.5$

$p_\theta(x \mid y=0) = \text{MVG} = N(x, \mu_0, \Sigma)$

$p_\theta(x \mid y=1) = \text{MVG} = N(x, \mu_1, \Sigma)$

$$p_\theta(y=c \mid x) = \frac{N(x_i, \mu_c, \Sigma) \times 0.5}{\sum_{c' \in \{0,1\}} N(x_i, \mu_{c'}, \Sigma) \times 0.5}$$

$$\frac{p_\theta(y=1 \mid x)}{p_\theta(y=0 \mid x)} = \frac{N(x_i, \mu_1, \Sigma)}{N(x_i, \mu_0, \Sigma)}$$

Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$$\arg\max_{\theta} \sum_{i=1}^{N} (1-y_i)\{-\log\{\exp(\theta^T x_i)+1\} + y_i(\theta^T x_i - \log\{\exp \theta^T x_i + 1\})$$

To predict $y=1$, $N(x_i, M_1, \Sigma) - N(x_i, M_0, \Sigma) > 0$

pdf MVG

$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\{-\tfrac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\}$$

$$-(x-M_1)^T\Sigma^{-1}(x-M_1) + (x-M_0)^T\Sigma^{-1}(x-M_0) > 0$$

$$\Rightarrow -x^T\Sigma^{-1}x + 2M_1^T\Sigma^{-1}x - M_1^T\Sigma^{-1}M_1$$
$$+ x^T\Sigma^{-1}x - 2M_0^T\Sigma^{-1}x + M_0^T\Sigma^{-1}M_0 > 0$$

$$\Rightarrow W^T x + b > 0$$
$$\theta^T x + b > 0$$

$$w = \Sigma^{-1}(M_1 - M_0)$$

$$b = \frac{1}{2}(M_0^T\Sigma^{-1}M_0 - M_1^T\Sigma^{-1}M_1)$$

LDA: Linear Discriminant Analysis

assump.

$$p_\theta(y=0) = p_\theta(y=1) = 0.5$$

$p_\theta(x|y=0) = MVG = N(x, M_0, \Sigma)$ $\quad p_\theta(x|y=1) = MVG = N(x, M_1, \Sigma)$

$$p_\theta(y=c|x) = \frac{N(x_i, M_c, \Sigma) \times 0.5}{\sum_{c' \in \{0,1\}} N(x_i, M_{c'}, \Sigma) \times 0.5}$$

$$\frac{p_\theta(y=1|x)}{p_\theta(y=0|x)} = \frac{N(x_i, M_1, \Sigma)}{N(x_i, M_0, \Sigma)}$$

Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$$\arg\max_{\theta} \sum_{i=1}^{N} (1-y_i)\{-\log\{\exp(\theta^T x_i)+1\} + y_i(\theta^T x_i - \log\{\exp\theta^T x_i + 1\})$$

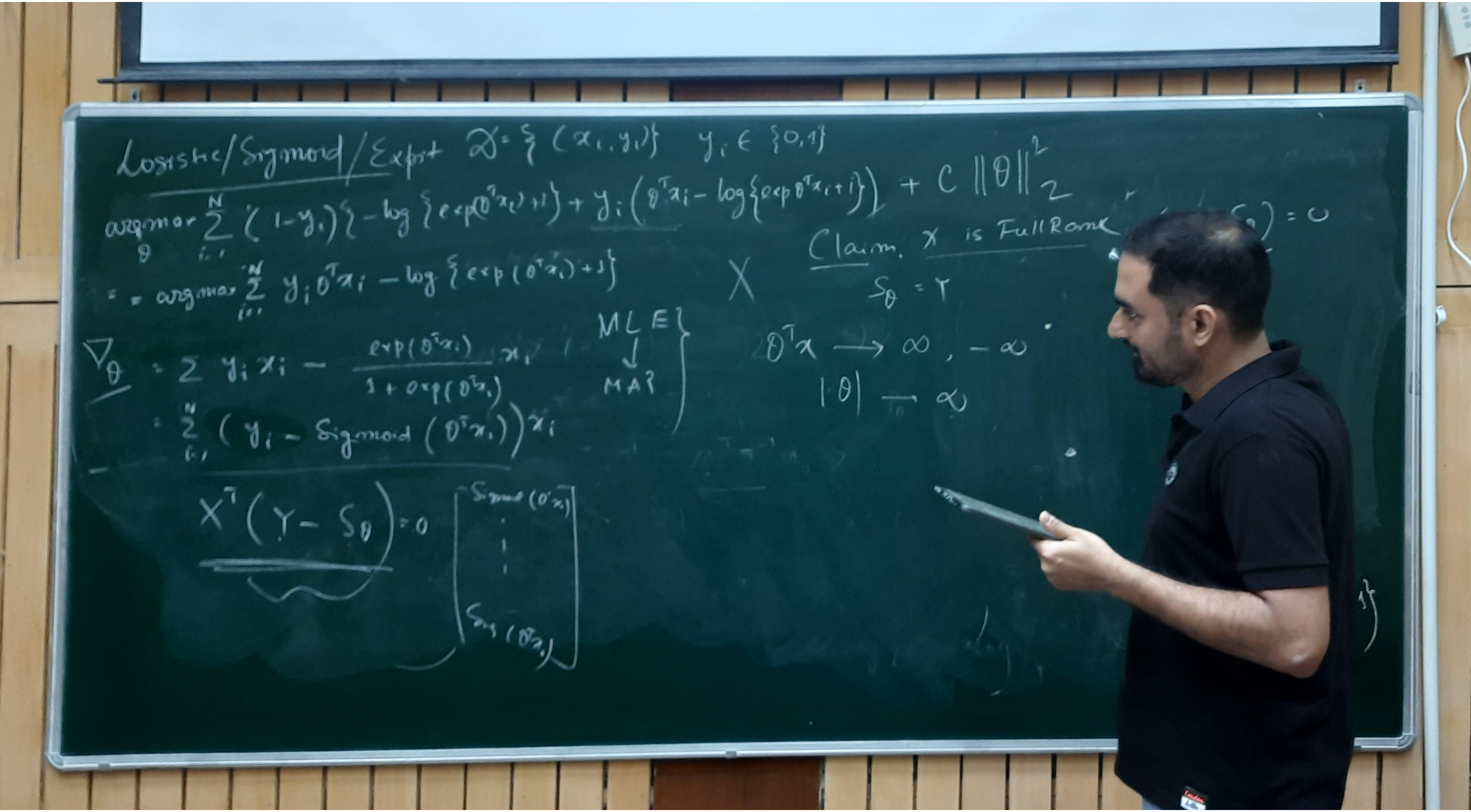
$$= \arg\max \sum_{i=1}^{N} y_i \theta^T x_i - \log\{\exp(\theta^T x_i)+1\}$$

$$\nabla_\theta = \sum y_i x_i - \frac{\exp(\theta^T x_i)}{1+\exp(\theta^T x_i)} x_i$$

$$= \sum_{i=1}^{N} (y_i - \text{Sigmoid}(\theta^T x_i)) x_i$$

$$X^T(Y - S_\theta)$$

$$\begin{bmatrix} \text{Sigmoid}(\theta^T x_1) \\ \vdots \\ \text{Sig}(\theta^T x_n) \end{bmatrix}$$



Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$
$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{N} (1-y_i)\{-\log\{\exp(\theta^T x_i)+1\}\} + y_i(\theta^T x_i - \log\{\exp \theta^T x_i + 1\}) + C\|\theta\|_2^2$
$= \operatorname{argmax} \sum_{i=1}^{N} y_i \theta^T x_i - \log\{\exp(\theta^T x_i)+1\}$
$\nabla_\theta = \sum y_i x_i - \frac{\exp(\theta^T x_i)}{1+\exp(\theta^T x_i)} x_i$
$= \sum_{i=1}^{N} (y_i - \text{Sigmoid}(\theta^T x_i)) x_i$
$X^T(Y - S_\theta) = 0$
$\begin{bmatrix} \text{Sigmoid}(\theta^T x_1) \\ \vdots \\ \text{Sig}(\theta^T x_n) \end{bmatrix}$
MLE
↓
MAP
X
Claim. X is Full Rank
$S_\theta = Y$
$\theta^T x \to \infty, -\infty$
$|\theta| \to \infty$

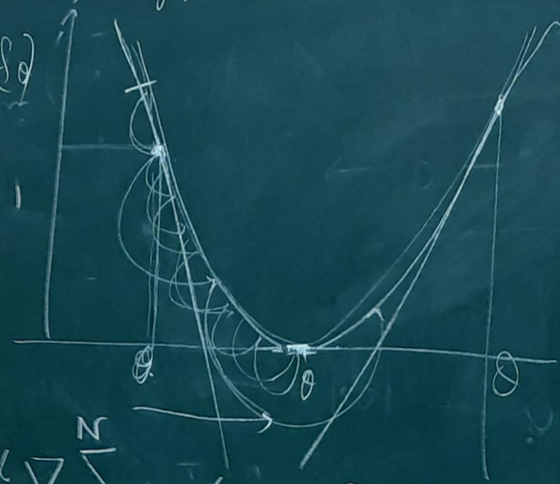
Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$\sum y_i \theta^T x_i - \log\{\exp(\theta^T x_i) + 1\}$ $l(\theta)$

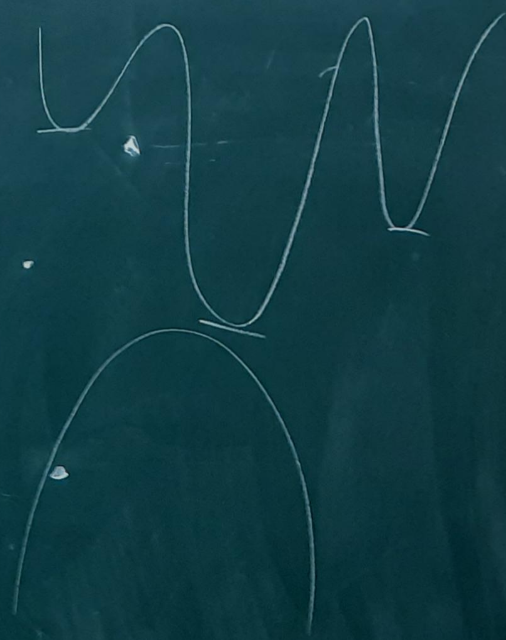
negative log likelihood → Convex

$\theta \leftarrow$ Initialization

for( )
if(not converged)
{
$\theta \leftarrow \theta - \alpha \nabla_\theta l(f_\theta(x))$ loss
}



$$\theta \leftarrow \theta - \alpha \nabla_\theta \sum_{i=1}^{N} l(f_\theta(x_i))$$



Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$\sum y_i \theta^T x_i - \log\{\exp(\theta^T x_i)+1\}$ $l(\theta)$

negative log likelihood → Convex

$\theta \leftarrow$ Initialization

for( )

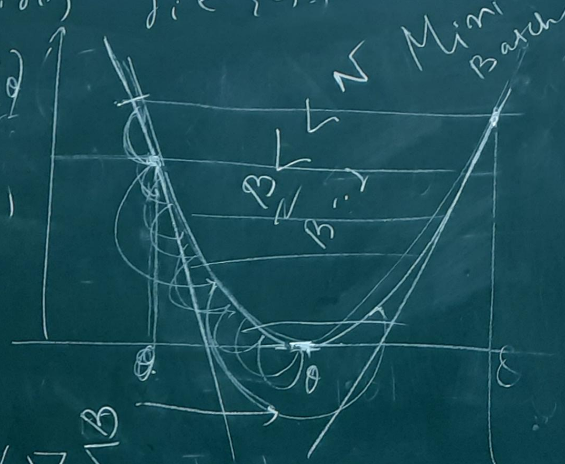
if (not converged)

{

$\theta \leftarrow \theta - \alpha \nabla_\theta l(f_\theta(x))$ loss

}

$\theta \leftarrow \theta - \alpha \nabla_\theta \sum_{i=1}^{B} l(f_\theta(x_i))$



Logistic/Sigmoid/Expit $\mathcal{D} = \{(x_i, y_i)\}$ $y_i \in \{0,1\}$

$\sum y_i \theta^T x_i - \log\{\exp(\theta^T x_i)+1\}$

negative log likelihood → Convex

$\theta \leftarrow$ Initialization

for(;)

if (not converged)

{

$\theta \leftarrow \theta - \alpha \nabla_\theta \, l(f_\theta(x))$ loss

}

}

$K > 2$

$\theta \,|_{m \times K}$

$\theta^T \big|_{K \times m} \, x \big|_{m \times 1} = \big|_{K \times 1}$

$\theta^T x$

$x \,|_m$

$$\frac{e^z}{\sum e^{z'}}$$

Soft MAX