

Logistic Regression

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

$$y_i = \text{label} \in \{0, 1, \dots, K-1\} \quad \text{Class } \begin{matrix} 0/n \\ 1/n \\ \vdots \\ K-1/n \end{matrix}$$

$$f_{\theta}(x) = \underline{y} \quad \begin{matrix} K=2 \\ \{0, 1\} \end{matrix}$$

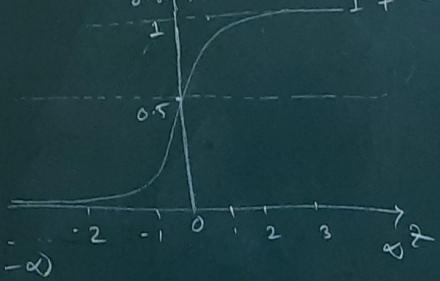
$$P_{\theta}(y=1|x) = 1 - P_{\theta}(y=0|x)$$

$$f_{\theta}(x) = \underline{\theta^T x} \quad \begin{matrix} -\infty \\ \rightarrow \infty \end{matrix}$$

How do we bound $f_{\theta}(x)$ so it produces values b/w 0 to 1?

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

$$\text{Sigmoid}(z) = \frac{1}{1 + e^{z}}$$



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

MLE of θ

$$\begin{aligned} & \arg \max \prod_{i=1}^N P_{\theta}(1|x_i) \underbrace{P_{\theta}(0|x_i)}_{y_i=0} \\ &= \arg \max \prod_{i=1}^N \underbrace{P_{\theta}(0|x_i)^{1-y_i}}_{y_i=0} \underbrace{P_{\theta}(1|x_i)^{y_i}}_{y_i=1} \end{aligned}$$

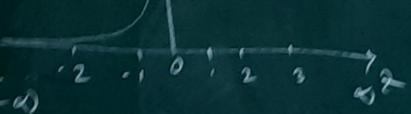
$$f_{\theta}(x) = \underline{\theta^T x} - \overline{\infty} \rightarrow \infty$$

- Unconstrained real numbers
How do we bound $f_{\theta}(x)$ s.t. it produces values between 0 to 1

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

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


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



MLE of θ

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

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

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

$$-\log \{ \exp \theta^\top x_i + 1 \} \quad \theta^\top x_i - \log \{ \exp \theta^\top x_i + 1 \}$$

$$\sigma(\text{sig}) = \text{sig}(1 - \text{sig})$$

$$f_\theta(x) = \underbrace{\theta^\top x}_{-\infty \rightarrow \infty} - \text{Unconstrained real numbers}$$

How do we bound $f_\theta(x)$ so it produces values between 0 to 1

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

$$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N (1-y_i) \{-\log \{\exp(\theta^T x_i + b)\} + y_i (\theta^T x_i - \log \{\exp(\theta^T x_i + b)\})\}$$

probabilistic

disseminative

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

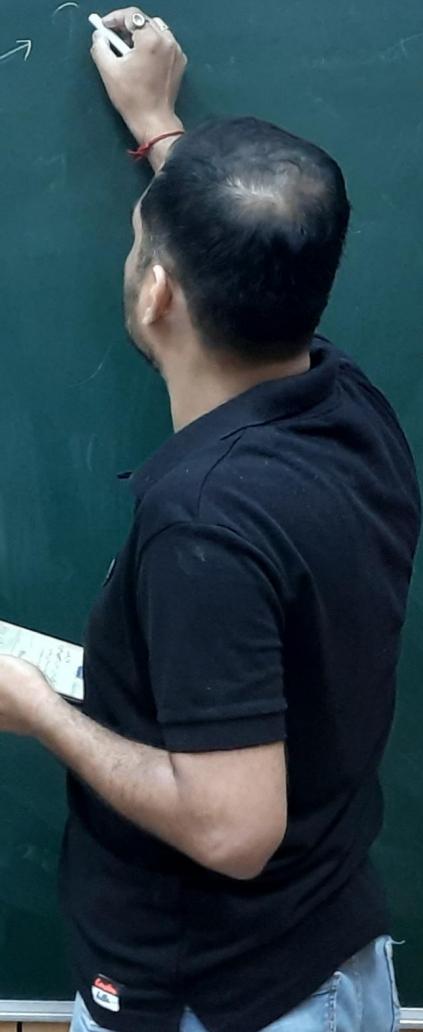
forged above —

Clam conditionals

Generative

Clam prior

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



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

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

DA : Linear Discriminant Analysis

$$P_0(y=0) = P_0(y=1) = 0.5$$

$$\begin{cases} P_0(x|y=0) & P_0(x|y=1) \\ = MVG & = MVG \\ = N(x, M_0, \Sigma) & = N(x, M_1, \Sigma) \end{cases}$$

$$P_0(y=c|x) = \frac{N(x_i, M_c, \Sigma) \times 0.5}{\sum_{c' \in \{0, 1\}} N(x_i, M_{c'}, \Sigma) \times 0.5}$$

$$\frac{P_0(y=1|x)}{P_0(y=0|x)} = \frac{N(x_i, M_1, \Sigma)}{N(x_i, M_0, \Sigma)}$$

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

$$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N \left[(1-y_i) \{-\log \{\exp(\theta^T x_i) + 1\} + y_i (\theta^T x_i - \log \{\exp(\theta^T x_i) + 1\})\} \right]$$

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

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

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

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

$$\Rightarrow w^T x + b > 0$$

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

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

LDA: Linear Discriminant Analysis

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

$$\begin{aligned} P_\theta(x|y=0) &= MVG \\ P_\theta(x|y=1) &= MVG \\ \sim N(x, M_0, \Sigma) &\quad \sim N(x, M_1, \Sigma) \end{aligned}$$

$$P_\theta(y=c|x) = N(x_i, M_c, \Sigma) \times 0.5$$

$$\frac{\sum_{c \in \{0,1\}} N(x_i | M_c, \Sigma) \times 0.5}{P_\theta(y=0|x)} = \frac{N(x_i, M_1, \Sigma)}{N(x_i, M_0, \Sigma) + 1}$$

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

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

$$J = \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N y_i \theta^T x_i - \log \{\exp(\theta^T x_i + 1)\}$$

$$\begin{aligned} \nabla_{\theta} J &= \sum y_i x_i - \frac{\exp(\theta^T x_i)}{1 + \exp(\theta^T x_i)} x_i^T \sum (1 - \frac{\exp(\theta^T x_i)}{1 + \exp(\theta^T x_i)}) x_i \\ &= \sum_{i=1}^N (y_i - \text{Sigmoid}(\theta^T x_i)) x_i \end{aligned}$$

$$X^T (Y - S_{\theta}) \underbrace{\begin{bmatrix} \text{Sigmoid}(\theta^T x_1) \\ \vdots \\ \text{Sigmoid}(\theta^T x_N) \end{bmatrix}}_{S_{\theta}}$$

Logistic/Sigmoid/Expt $\hat{y} = \sigma(\mathbf{x}_i, \mathbf{y}_i)$ $y_i \in \{0, 1\}$

$$\underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N (1-y_i) \{-\log \{\exp(\theta^T \mathbf{x}_i + b)\} + y_i (\theta^T \mathbf{x}_i - \log \{\exp(\theta^T \mathbf{x}_i + b)\}) + C \|\theta\|_2^2\}$$

$$= \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^N y_i \theta^T \mathbf{x}_i - \log \{\exp(\theta^T \mathbf{x}_i + b)\} \quad \times \quad \begin{array}{l} \text{Claim: } \mathbf{x} \text{ is Full Rank} \\ S_\theta = \mathbf{I} \end{array}$$

$$\begin{aligned} \nabla_{\theta} &= \sum_{i=1}^N y_i \mathbf{x}_i - \frac{\exp(\theta^T \mathbf{x}_i)}{1 + \exp(\theta^T \mathbf{x}_i)} \mathbf{x}_i \\ &= \sum_{i=1}^N (y_i - \text{Sigmoid}(\theta^T \mathbf{x}_i)) \mathbf{x}_i \end{aligned} \quad \left. \begin{array}{l} \text{ML} \\ \text{MAP} \end{array} \right\} \quad \begin{array}{l} \theta^T \mathbf{x} \rightarrow \infty, -\infty \\ |\theta| \rightarrow \infty \end{array}$$

$$\mathbf{x}^T (\mathbf{y} - \mathbf{S}_{\theta}) = 0 \quad \left[\begin{array}{c} \text{Sigmoid}(\theta^T \mathbf{x}_1) \\ \vdots \\ \text{Sigmoid}(\theta^T \mathbf{x}_N) \end{array} \right]$$



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

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

negative log likelihood \rightarrow convex

θ ← Initialization

for
if (not converged)

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

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



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

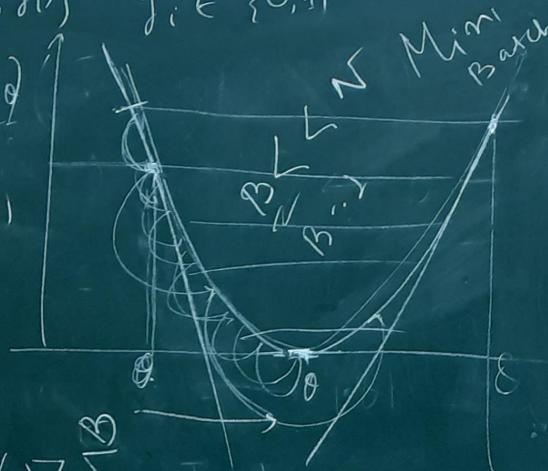
$$\sum y_i \theta^T x_i - \log \left\{ \exp(\theta^T x_i) + 1 \right\} l(\theta)$$

negative log likelihood \rightarrow Convex

$\theta \leftarrow$ Initialization
for
if (not converged)

$$\begin{cases} \theta \leftarrow \theta - \alpha \nabla_{\theta} l(f_{\theta}(x)) \\ \text{loss} \end{cases}$$

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



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

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

negative log likelihood \rightarrow Convex

$\theta \leftarrow$ Initialization
for
if (not converged)

$$\theta \leftarrow \theta - \alpha \nabla_{\theta} \ell(f_{\theta}(x))$$

$$K > 2$$

$$\theta | m \times K$$

$$\theta^T x | m \times 1 | K \times 1$$

$$\theta^T x$$

$$x^T m$$

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

Softmax.

