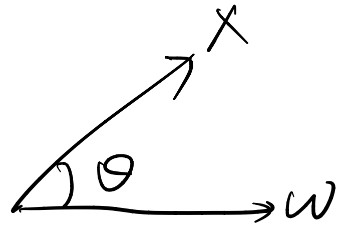
25. 09. 22.

Geometric Interpretation of LR.



$X, W \in \mathbb{R}^D$

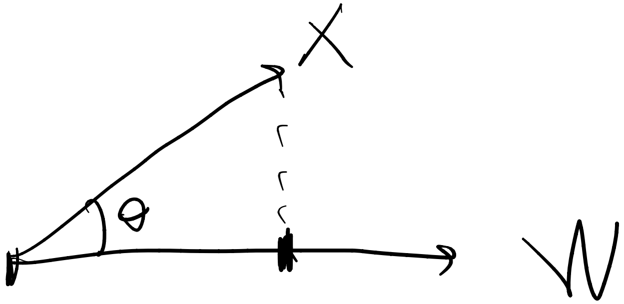
$$\boxed{W^T X = \|W\| \|X\| \cos\theta}$$

$$W^T X = w_1 x_1 + U_2 x_2 + \cdots + w_D x_D = \|W\| \|X\| \cos\theta$$

$$\|X - w\|^2 = (X - w)^T (X - w) = X^T X + w^T w - 2 w^T X$$

$$= \boxed{\|X\|^2 + \|w\|^2 - 2 w^T X}$$

$$\|X - w\|^2 = \boxed{\|X\|^2 + \|w\|^2 - 2 \|X\| \|W\| \cos\theta}$$



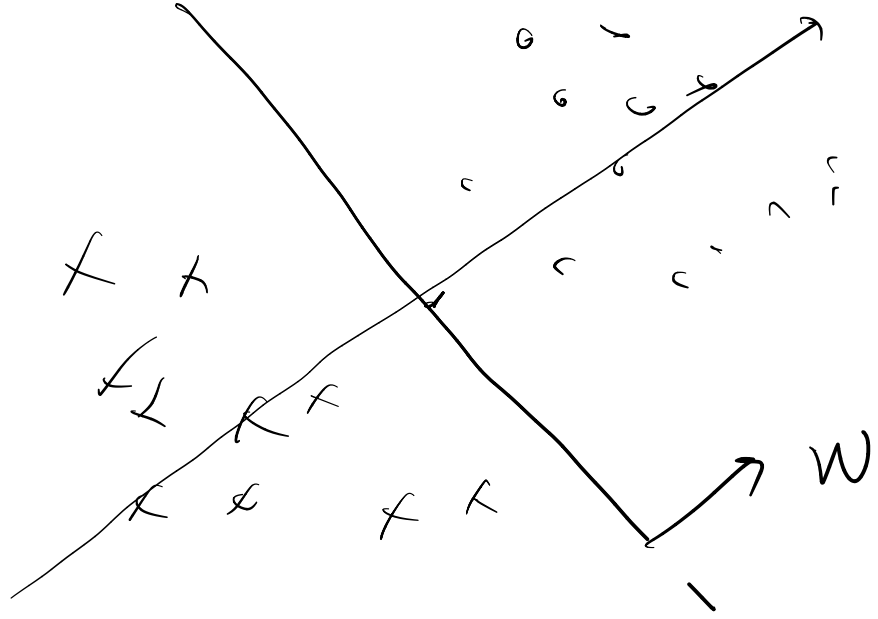
$$\|x\|\cos\theta = \cancel{\|x\|}\cdot\frac{w^Tx}{\|w\|\cancel{\|x\|}}$$

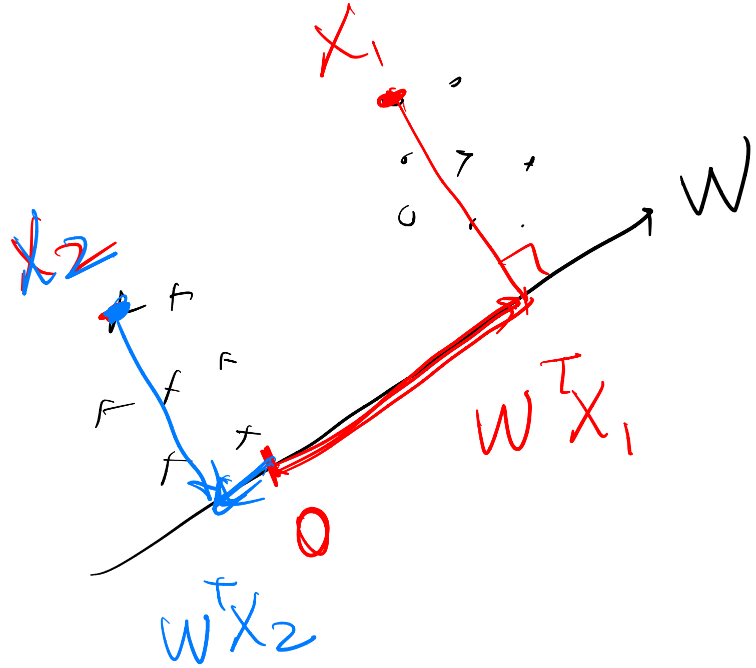
$$= \frac{w^Tx}{\|w\|}$$

: Length of projected vector.

· Projected vector:

$$\frac{w^Tx}{\|w\|}\cdot\frac{w}{\|w\|}$$





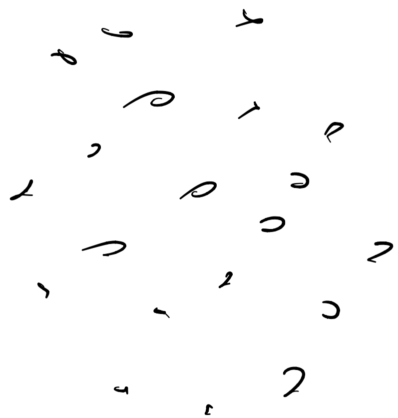
$\|w\| = 1$

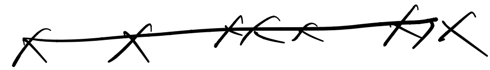
$$\begin{cases} w^T x_i \geq b \rightarrow y_i = 1 \\ w^T x_i < b \rightarrow y_i = 2 \end{cases}$$

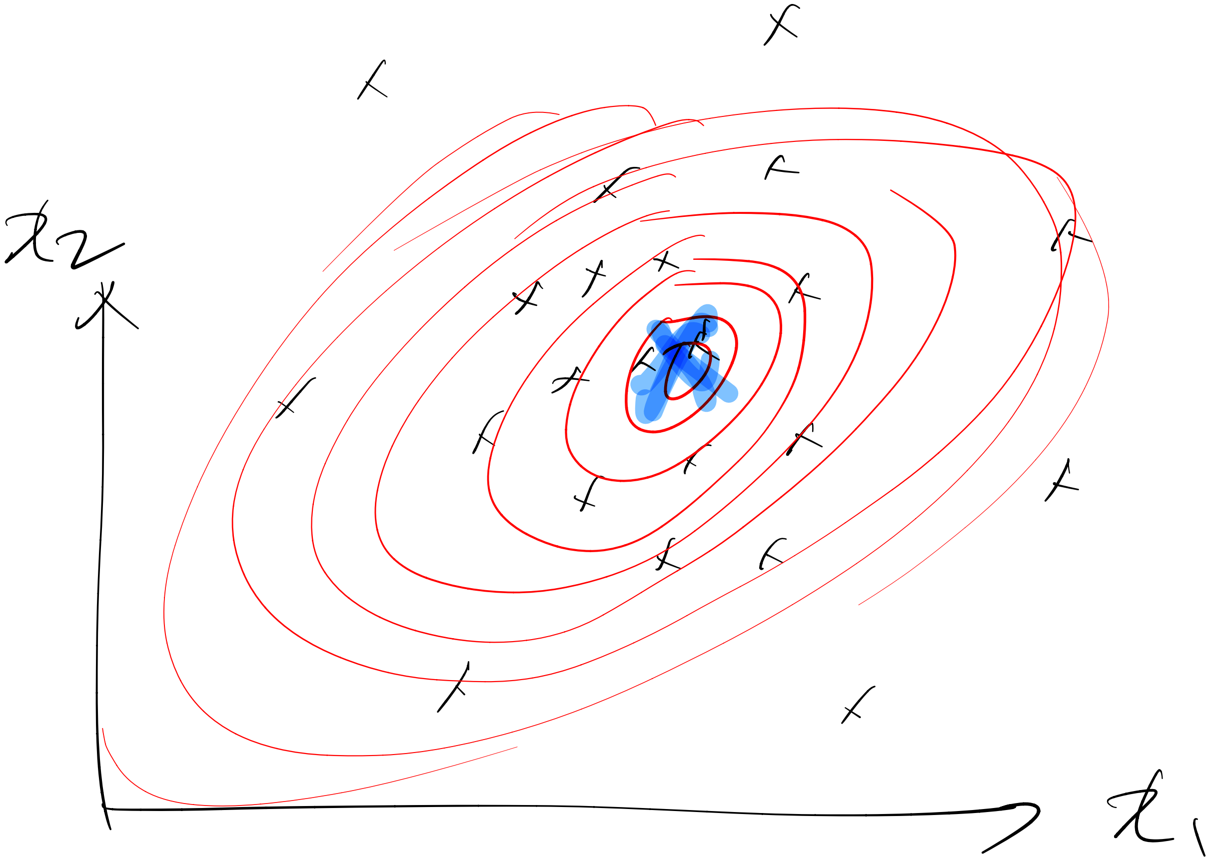
$$\sum_{i=1}^{N} \frac{1}{2} \left\| \underline{f(x_i ; w)} - \underline{y_i} \right\|^2 \qquad D = \{x_i, y_i\}$$

MSE loss

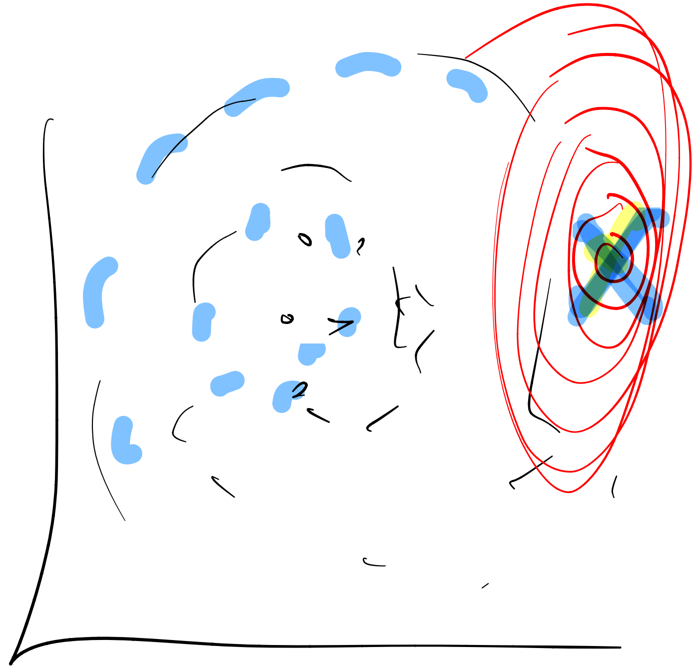
Likelihood

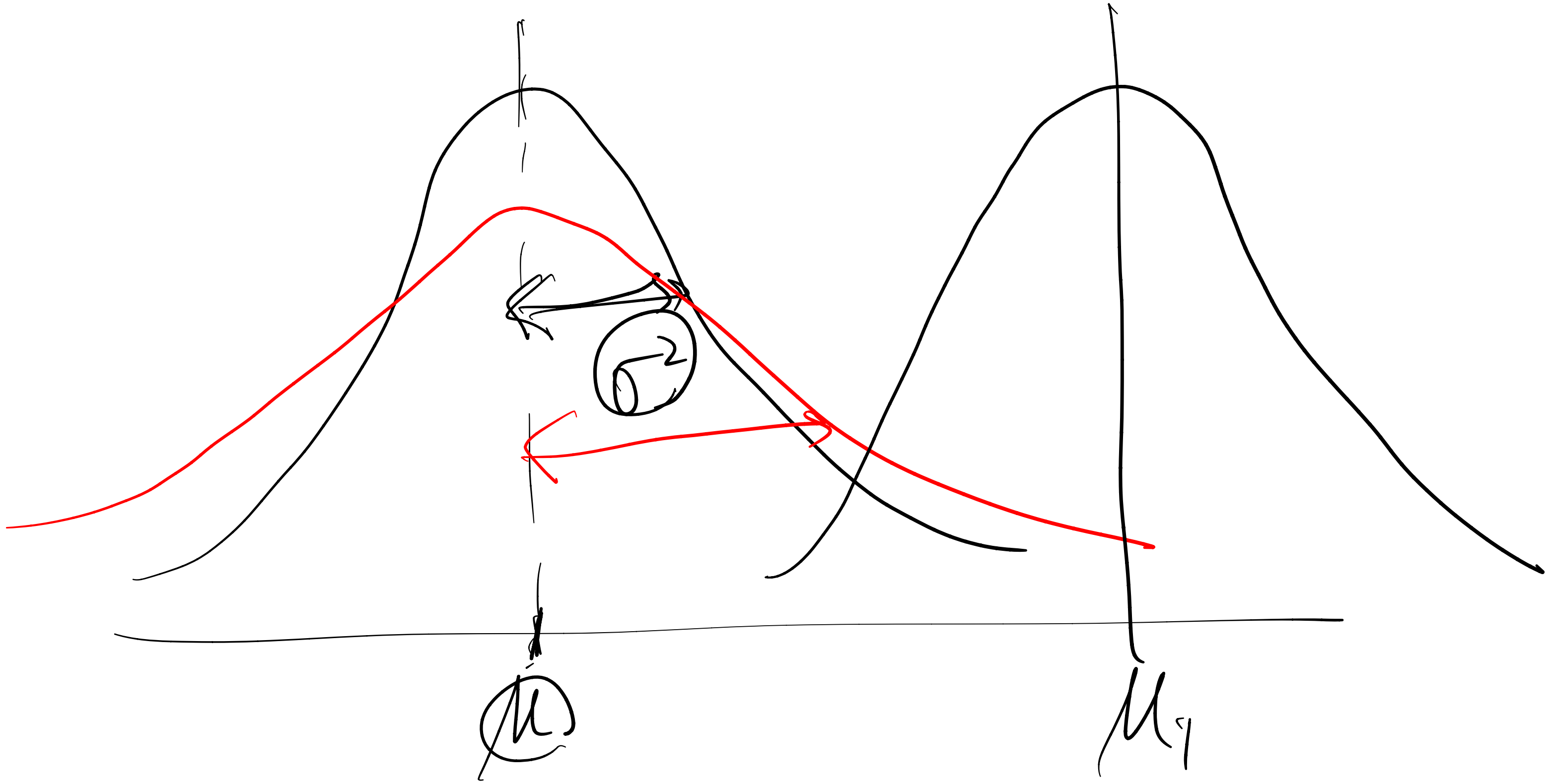






x_2
x_1





σ^2
μ
μ_1

Logistic Regression Revisited.

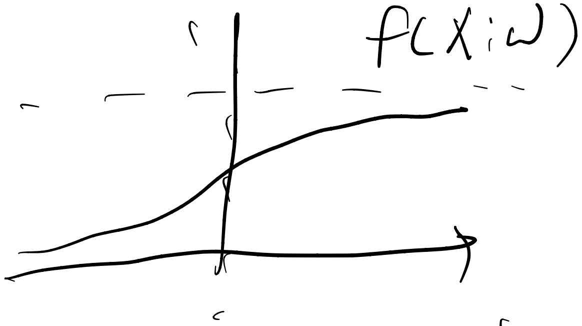
$y_i \in \{0, 1\}$

$x_i \in \mathbb{R}^D$

$i = 1, \cdots, N$

$$f(x; w) = \frac{1}{1 + \exp(-w^T x)}$$

$$\max L(w) = \prod_{i=1}^{N} f(x_i)^{y_i} (1 - f(x_i))^{(1-y_i)}$$



$f(x_i)$: Prob. of choosing $y_i = 1$

$1 - f(x_i)$: Prob. " " $y_i = 0$.

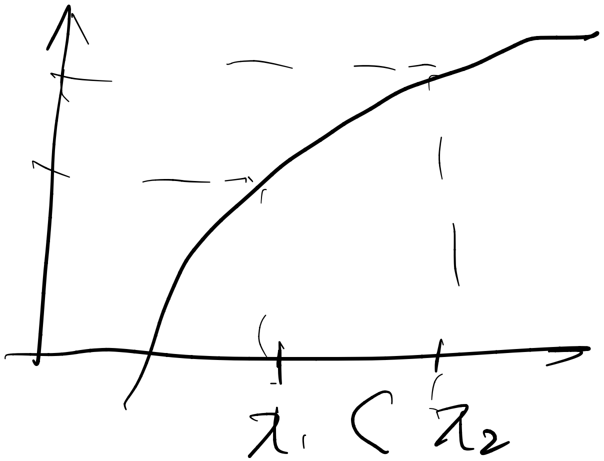
$L(w)$: Prob. that $f(x)$ classifies all data correctly.

$$\boxed{\max_{w} L(w) = \prod_{i=1}^{N} f(x_i)^{y_i} \{1 - f(x_i)\}^{1-y_i}}$$

$$\boxed{\max_{w} \ln L(w) = \sum_{i=1}^{N} y_i \ln f(x_i) + (1-y_i) \ln (1 - f(x_i))}$$

$$w^* = \arg\max_{w} L(w) = \arg\max_{w} \ln L(w)$$

$\because$ ln is a monotonically Increasing function.



$$y\ln f + (1-y)\ln(1-f) \to ①$$

$$\frac{d①}{dw} = y\cdot\frac{1}{f}\cdot\frac{df}{dw} + (1-y)\cdot\frac{1}{1-f}\cdot\frac{d(1-f)}{dw}$$

$$\boxed{\frac{df}{dw} = f(1-f)X}$$

$$= y\cdot\frac{1}{\cancel{f}}\cdot\cancel{f}(1-f)X + (1-y)\cdot\frac{1}{\cancel{1-f}}\cdot\left(-f\cancel{(1-f)}X\right)$$

$$= y(1-f)X - (1-y)f(X) = (y-f)X$$

$$\frac{\partial}{dw} \ln L(w) = \sum_{i=1}^{N} (y_i - f(x_i)) x_i .$$

Update rule :

$$w_{t+1} = w_t + \eta \sum_{i=1}^{N} (y_i - f(x_i)) x_i$$