25. 9. 15.

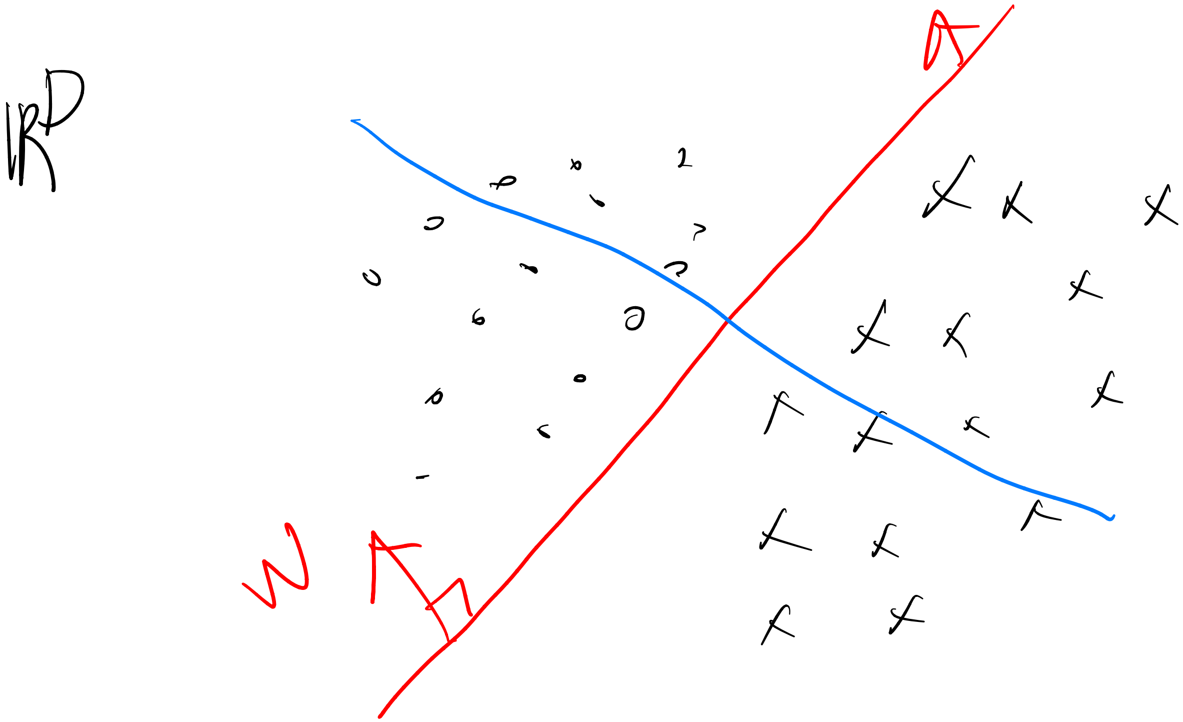
Numerical Optimization
Gradient descent.

Last time. $f: \mathbb{R}^p \to \mathbb{R}$ $\quad x \mapsto y.$

$$\underline{\frac{df}{dx} \in \mathbb{R}^p.} \qquad \left[\frac{df}{dx}\right]_i = \frac{\partial f}{\partial x_i}.$$

Linear Classifier.

$D = \{x_i, y_i\}_{i=1}^{N}$ is given. $x_i \in \mathbb{R}^p$, $y_i \in \{0,1\}$.



$W^T x_i - b \geq 0$

$\rightarrow y_i = 1$

$W^T x_i - b < 0$

$\rightarrow y_i = 0$.

W : normal vector.

$W \in \mathbb{R}^p$.

$$L(w) = \frac{1}{2} \sum_{i=1}^{N} \| f(\underset{\sim}{x_i}; w) - y_i \|^2$$

$$w^* = \underset{w}{\operatorname{argmin}} \, L(w)$$

Learning (optimization).

$$L(w) = \frac{1}{2}\sum_{i=1}^{N} \| f(x_i; w) - y_i \|^2$$

$$f(x; w) = w^T x \qquad (b=0)$$

$$L(w) = \frac{1}{2}\sum_{i=1}^{N} \| w^T x_i - y_i \|^2$$

$$= \frac{1}{2} \| X^T w - y \|^2$$

$$= \frac{1}{2} (X^T w - y)^T (X^T w - y)$$

$$\frac{dL}{dw} = \frac{1}{2} \cdot 2 \cdot X (X^T w - y)$$

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

$X = \begin{pmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_N \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{D \times N}$

$y \in \mathbb{R}^N$

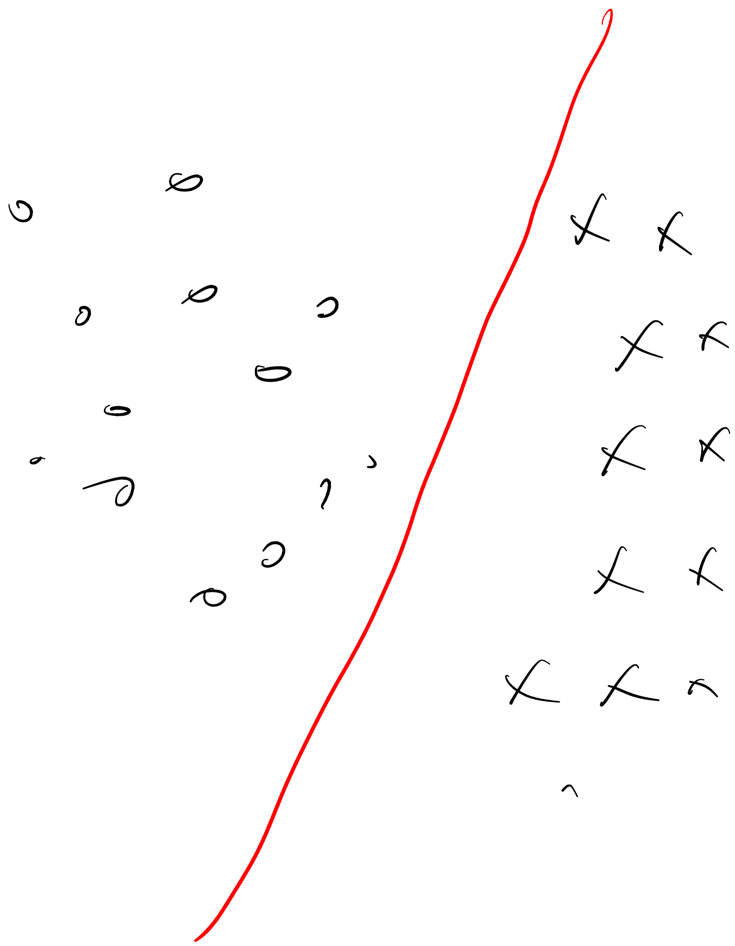
$$\boxed{\left.\frac{dL}{dw}\right|_{w=w^*} = 0}$$

- $\frac{dL}{dw} = \frac{1}{\cancel{2}} \cdot \cancel{2} \cdot X (X^T w - y) = 0$

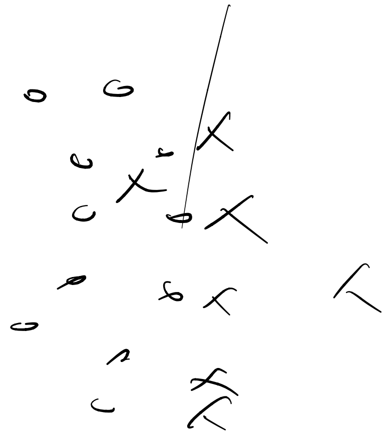
$$X X^T \underset{\sim}{w} - Xy = 0$$

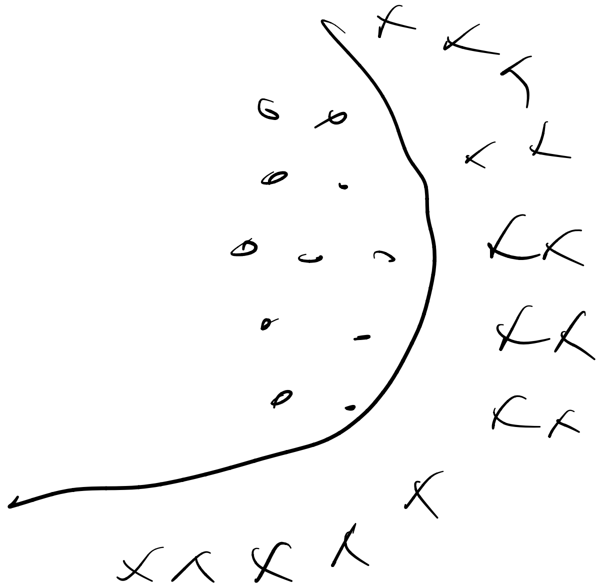
$$X X^T w = Xy$$

$$w^* = (X X^T)^{-1} X y$$



linear separable.





$$= \frac{dL}{dw} = w \cdot \sum_{i=1}^{N} (w^T x_i - y_i)$$

과연
방정식
풀어..

Analytical Method.

$$\frac{dL}{dw} = 0. \Rightarrow w = L(w)$$

Numerical Method.

$$w_{t+1} = w_t - \eta \cdot \frac{dL}{dw}\bigg|_{w=w_t}$$