Dataset = $\{(x_1,y_1), \dots, (x_n,y_n)\}$ $\quad x_i \in \mathbb{R}^d$, $y_i \in \{+1,-1\}$

Goal: $h : \mathbb{R}^d \to \{\pm 1\}$

Performance measure

$$\sum_{i=1}^{n} \mathbb{1}(h(x_i) \neq y_i)$$

$\mathbb{1}(z) = \begin{cases} 1 & \text{if } z \text{ is true} \\ 0 & \text{o/w} \end{cases}$

$$\min_{h \in \mathcal{H}_{linear}} \sum_{i=1}^{n} \mathbb{1}(h(x_i) \neq y_i)$$

$$\equiv \min_{w \in \mathbb{R}^d} \sum_{i=1}^{n} \mathbb{1}(\text{sign}(w^T x_i) \neq y_i)$$

$\hookrightarrow$ NP-HARD

$x \in \mathbb{R}^d$

$x' = [x \; 1] \in \mathbb{R}^{d+1}$

$w' = [\underbrace{w_1 \cdots w_d}_{w} \; \underbrace{w_{d+1}}_{b}]$

$w'^T x' = w^T x + b$

LOSS-FUNCTION VIEW

$(\underset{\mathbb{R}^d}{x}, \underset{\{\pm 1\}}{y})$, $\quad h : \mathbb{R}^d \to \{\pm 1\}$

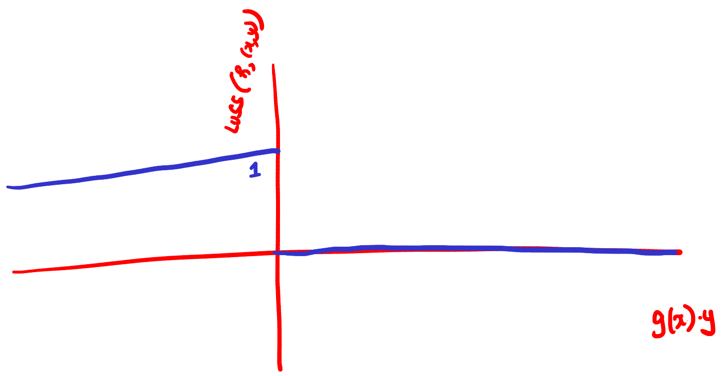
$h(x) = \text{sign}(w^T x)$

$$\mathbb{1}(h(x) \neq y) = \mathbb{1}((\underset{\geq 0}{w^T x}) \cdot \underset{+1}{y} < 0)$$

$\leq 0 \quad -1$

$> 0 \quad -1$

$< 0 \quad +1$



$h(x) = \text{sign}(g(x))$

$$\sum_{i=1}^{n} \mathbb{1}(g(x_i) y_i < 0)$$

ALG 1: Using Regression for classification $\quad h(x) = \text{sign}(g(x))$

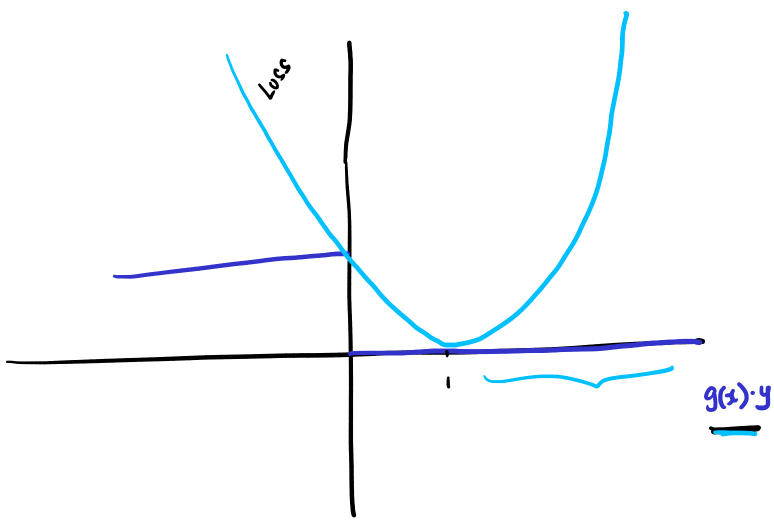
$$\text{Loss}(g, (x,y)) = (g(x) - \underset{\{\pm 1\}}{y})^2 = (g(x))^2 + y^2 - 2g(x)\cdot y$$

$$= (g(x))^2 + 1 - 2g(x)\cdot y \quad -①$$

$$= \boxed{(g(x)y - 1)^2} \qquad = (g(x)\cdot y)^2 + 1 - 2g(x)\cdot y$$

$= (g(x))^2 + 1 - 2g(x)\cdot y$ — ②

① = ② .



- → 0-1 loss
- → $(g(x)\cdot y - 1)^2$ → SQUARED LOSS

SUPPORT VECTOR MACHINES

$$\min_{w,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t} \quad (w^T x_i)y_i + \xi_i \geq 1$$

$$\xi_i \geq 0$$

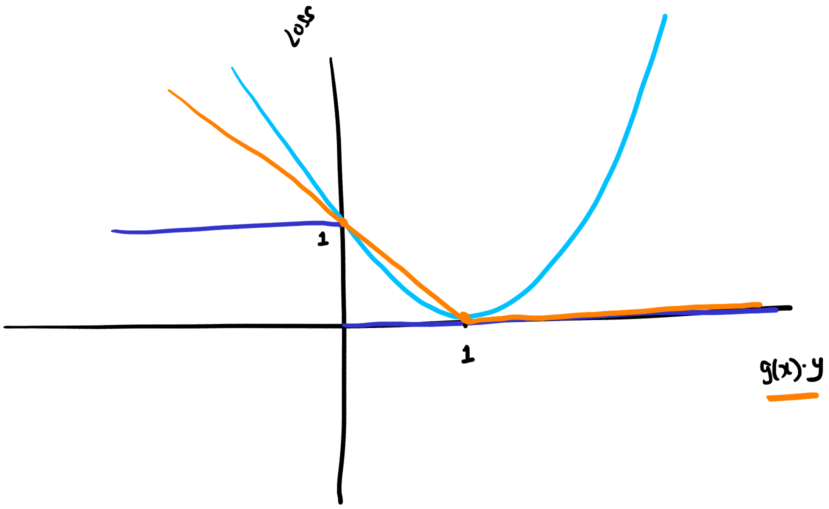
$$\min_{w,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t} \quad \begin{cases} \xi_i \geq 1 - (w^T x_i)y_i \\ \xi_i \geq 0 \end{cases}$$

$$\xi_i \geq \max(0, 1-(w^T x_i)y_i)$$

$$= \quad \min_{w\in\mathbb{R}^d} \quad \underbrace{\overset{\text{Regularizer}}{\frac{1}{2}\|w\|^2}}_{\text{Model dependent}} + \underbrace{C\sum_{i=1}^{n}\max(0, 1-\underbrace{(w^T x_i)}_{g(x_i)}y_i)}_{\text{Data dependent Loss}}$$

REGULARIZATION + HINGE LOSS



- 0-1 loss
- Squared loss $(g(x)-1)^2$
- $\max(0, 1-g(x)\cdot y)$ HINGE LOSS

LOGISTIC REGRESSION

$\sigma(a) = \frac{1}{1+e^{-a}}$

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

$$\max_{w} \prod_{i=1}^{n} \left(\sigma(w^T x_i)\right)^{z_i} \left(1 - \sigma(w^T x_i)\right)^{(1-z_i)}$$

$z_i = 1$ if $y_i = +1$

$z_i = 0$ if $y_i = -1$

$$\max_{w} \sum_{i=1}^{n} z_i \log\left(\sigma(w^T x_i)\right) + (1-z_i) \log\left(1 - \sigma(w^T x_i)\right)$$

$$\equiv \min_{w} \sum_{i=1}^{n} \left[-z_i \log\left(\sigma(w^T x_i)\right) + (z_i - 1) \log\left(1 - \sigma(w^T x_i)\right)\right]$$

Loss for a single point $z_i = 1$ $(y_i = +1)$

$$= -\log\left(\sigma(w^T x_i)\right) = -\log\left(\frac{1}{1+e^{-w^T x_i}}\right)$$

$$= \log\left(1 + e^{-w^T x_i}\right) = \boxed{\log\left(1 + e^{-(w^T x_i) y_i}\right)}$$

Loss for a single point $z_i = 0$ $(y_i = -1)$

$$-\log\left(1 - \sigma(w^T x_i)\right)$$

$$= -\log\left(1 - \frac{1}{1+e^{-w^T x_i}}\right)$$

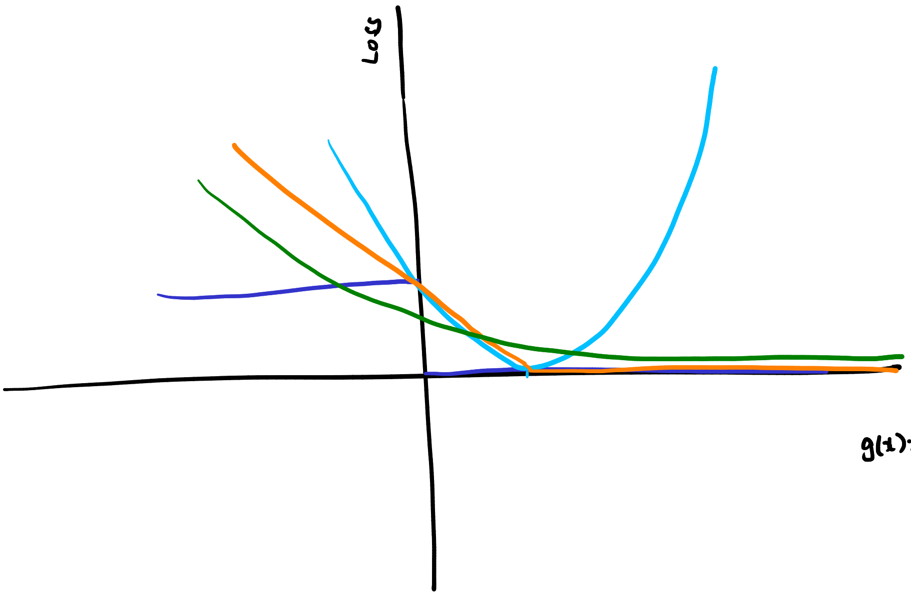
$$= -\log\left(\frac{e^{-w^T x_i} / e^{-w^T x_i}}{(1+e^{-w^T x_i}) / e^{-w^T x_i}}\right)$$

$$= -\log\left(\frac{1}{e^{w^T x_i} + 1}\right)$$

$$= \log\left(1 + e^{w^T x_i}\right)$$

$$= \boxed{\log\left(1 + e^{-w^T x_i y_i}\right)}$$

$$= \min_{w} \sum_{i=1}^{n} \log\left(1 + e^{\underbrace{-w^T x_i y_i}_{g(x_i) y_i}}\right)$$



CONCLUSIONS

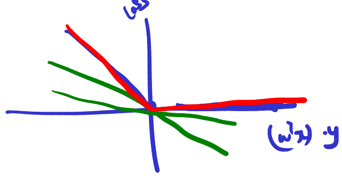
- 0-1 loss is NP-hard to minimize
- Different algorithms use different "surrogate" loss
- Surrogates are convex and hence easy to minimize.

PERCEPTRON

$$w_{t+1} = w_t + \underline{x_t y_t}$$

HINGE LOSS

$$\ell_{\text{mod-hinge}}(w, (x,y)) = \underline{\max(0, -(w^T x) y)}$$



$$\nabla_w \ell_{hinge} = \begin{cases} -xy & (w^T x)y < 0 \\ 0 & (w^T x)y > 0 \\ [-1, 0]\, xy & (w^T x)y = 0 \end{cases}$$

mistake

↳ chooses $-x \cdot y$ when mistake, 0 o/w.

$$w_{t+1} = w_t - \eta_t \underbrace{\nabla_w \, loss(w_t)}$$

$\eta_t = 1$

$$= w_t - (-x_t y_t)$$

- Perceptron can be interpreted as S.G.D with modified hinge loss with step size = 1

BOOSTING

- $Loss(h, (x, y)) = e^{-y\, h(x)}$

↳ Exponential loss.