2.1

$$y = [1 \;\; x \;\; x^2 \; \cdots \; x^M] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_M \end{bmatrix}$$

Root Mean Squared Error

$$E_{RMS} = \sqrt{\frac{f(\beta)}{n}}$$

$M \Uparrow \rightsquigarrow$ overfitting

Regularization

$$J[f] = f(\beta) + \lambda \overset{\Downarrow}{\sum_j |\beta_j|^q}$$

$$= \|y - X\beta\|^2 + \lambda \|\beta\|^2$$

$$= (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta$$

$$\nabla_\beta J[f] = \nabla_\beta \left(\cancel{y^T y} - \underbrace{y^T X \beta}_{-X^T y} - \underbrace{\beta^T X^T y}_{-X^T y} + \underbrace{\beta^T X^T X \beta}_{2X^T X\beta} \right) + \underbrace{\nabla_\beta \lambda \beta^T \beta}_{2\lambda\beta} .$$

$$= 0 \implies \begin{bmatrix} X^T(Y - X\beta) = \lambda\beta \\ \Downarrow \qquad\quad \Downarrow \end{bmatrix}$$

$x^T\beta \rightsquigarrow \phi(x)^T\beta$

low dimension — high dimension

<Kernel>

$$k(x_1, x_2) = \phi(x_1)^T \phi(x_2)$$

"inner-product"

$\cdot\ J[f] = \|y - \phi(X)^T\beta\|^2 + \lambda\|\beta\|^2$.

$$= y^Ty - y^T\phi(X)^T\beta - \beta^T\phi(X)y + \beta^T\phi(X)\phi(X)^T\beta + \lambda\beta^T\beta$$

$\cdot\ \nabla_\beta J[f] = 0 - \phi(X)y - \phi(X)y + 2\phi(X)\phi(X)^T\beta + 2\lambda\beta$

$$= -2\phi(X)y + 2\phi(X)\phi(X)^T\beta + 2\lambda\beta$$

$$= 0 \quad \text{(극값)}$$

$\cdot\ \beta^* = (\lambda I + \phi(X)\phi(X)^T)^{-1}\phi(X)y$

$$= \phi(X)(\lambda I + \phi(X)^T\phi(X))^{-1}y$$

$$= \phi(X)\underbrace{\boxed{(\lambda I + K(X,X))^{-1}y}}_{\alpha^*}$$

$\cdot\ f^*(x) = \phi(x)^T\beta^*$

$$= K(x, X)\cdot\alpha^*$$

example let $k(x,z) = (x^T z)^2$

$$= \langle \underbrace{(x_1^2,\ x_2^2,\ \sqrt{2}\, x_1 x_2)}_{\phi(x)},\ \underbrace{(z_1^2,\ z_2^2,\ \sqrt{2}\, z_1 z_2)}_{\phi(z)} \rangle$$

let $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $X_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ $Y = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$

then $\beta^* = \phi(X) \cdot \alpha^*$

$$= \phi(X)\,(\lambda I + k(X,X))^{-1}\, y.$$

$$k(X,X) = \begin{bmatrix} (1\cdot1+1\cdot1)^2 & (2\cdot1+0\cdot1)^2 \\ (1\cdot2+1\cdot0)^2 & (2\cdot2+0\cdot0)^2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 16 \end{bmatrix}$$

$\lambda I + k(X,X) = \begin{bmatrix} 5 & 4 \\ 4 & 17 \end{bmatrix}$, $\alpha^* = \begin{bmatrix} 5 & 4 \\ 4 & 17 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \frac{1}{69} \cdot \begin{bmatrix} 43 \\ -2 \end{bmatrix}$

$(\lambda = 1)$

$$f^*(x) = k(x, X) \cdot \alpha^*$$

$$= [1, 4] \cdot \frac{1}{69} \begin{bmatrix} 43 \\ -2 \end{bmatrix}$$

$x^T = [1, 0]$

cf. $k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right)$

$$k_2(x,z) = \frac{k_1(x,z)}{\sqrt{k_1(x,x)}\sqrt{k_1(z,z)}}$$

(sigmoid)

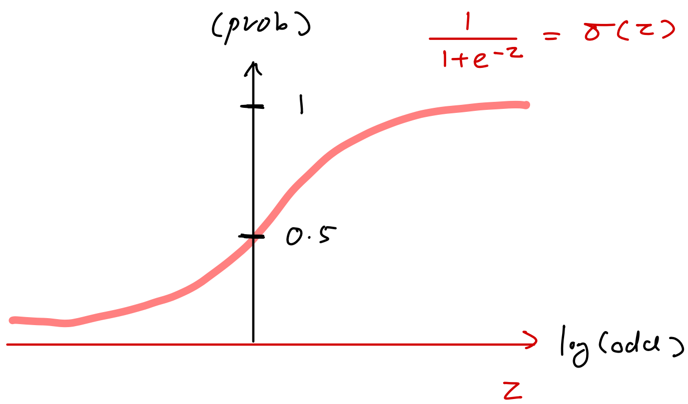
$$\begin{cases} \text{probability} & p = \dfrac{e^z}{1+e^z} \in [0, 1] \\ \text{odds} & o = e^z \in [0, \infty) \end{cases}$$

$$o = \frac{p}{1-p} \qquad p = \frac{o}{1+o}$$

$\downarrow$

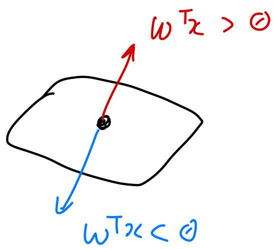
log (odds) z

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



Logistic Regression (MLE)

. hyperplane $w^T x = 0$



$$\begin{cases} \text{정답.} & y_i \in \{0, 1\} \\ \text{예측} & \sigma(\underbrace{w^T x}_{\log(\text{odd})}) \in [0, 1] \end{cases}$$

- $P(y_i = 1 \mid x_i, w) = \sigma(w^T x_i)$

- $$\mathcal{L} = \prod_i P(y_i \mid x_i, w)$$

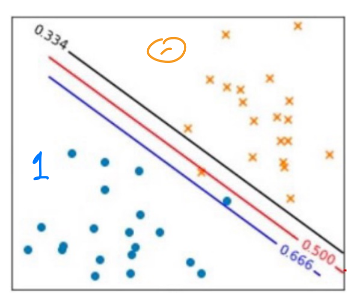
$$= \prod_i P(y_i = 1 \mid x_i, w)^{y_i} \, P(y_i = 0 \mid x_i, w)^{1 - y_i}$$

$$-\log \mathcal{L} = -\left[\sum_i y_i \log \frac{1}{1 + e^{-w^T x_i}} + \sum_i (1 - y_i) \log \frac{e^{-w^T x_i}}{1 + e^{-w^T x_i}} \right]$$

$$= \sum_i \log (1 + e^{-w^T x_i}) + \sum_i (1 - y_i) \, w^T x_i$$

$$\nabla_w (-\log \mathcal{L}) = \sum_i \frac{e^{-w^T x_i}}{1 + e^{-w^T x_i}} \cdot (-x_i) + \sum_i (1 - y_i) \, x_i$$

$$= \sum_i \left[\underbrace{\sigma(w^T x_i)}_{\text{예측 확률}} - \underbrace{y_i}_{\text{실제}} \right] \underbrace{x_i}_{\text{방향}}$$



빨간 선 $w^T x \geq 0$ prob ≥ 50%

파란선 $w^T x \geq$? prob ≥ 66%

claim $J = -\log \mathcal{L}$ 일 때, J는 convex 하고 global min을 가진다.

$J = -\log \mathcal{L} = \sum l_i$

$\nabla_w l_i = \left(\sigma(w^T x_i) - y_i\right) x_i$

$\nabla_w J = \sum_i \left(\sigma(w^T x_i) - y_i\right) x_i$

Note. 헤세 행렬은 대칭행렬이다.

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$$

- $\nabla_w (\nabla_w l_i) = \left[\nabla_w \left(\sigma(w^T x_i) - y_i\right)\right] \cdot x_i^T$

$$= \sigma'(w^T x_i) \cdot x_i x_i^T$$

Note. $\sigma' = \sigma(1-\sigma)$
≥ 0

- $\nabla_w (\nabla_w J) = \sum_i \sigma'_i \, x_i x_i^T$

$$= \begin{array}{ccc} X^T & S & X \\ \begin{bmatrix} | & & \\ x_1 & \cdots & \\ | & & \end{bmatrix} & \begin{bmatrix} \sigma' & & \\ & \ddots & \\ & & \sigma' \end{bmatrix} & \begin{bmatrix} - & x_1 & - \\ & \ddots & \\ & & \end{bmatrix} \end{array}$$

- for $\forall v \in \mathbb{R}^d$ $\quad v^T (\nabla_w^2 J) v = v^T X^T S X v$

$$= (Xv)^T S (Xv)$$

$$= \sum_i \underbrace{\sigma'_i}_{\geq 0} \underbrace{\left((Xv)_i\right)^2}_{\geq 0} \geq 0$$

positive semi-definite!
$\langle v, Hv \rangle \geq 0$

$\Longleftrightarrow$ J convex
: global min

$$\text{multiclass classification}: \quad p(y=k \mid x, W) = \frac{e^{w_k^T x}}{\sum_j e^{w_j^T x}}$$

assume $x = \begin{bmatrix} 5.1 \\ 3.5 \\ 1.4 \\ 0.2 \end{bmatrix}$ $w_1 = \begin{bmatrix} 0.1 \\ 0.2 \\ 0.3 \\ 0.4 \end{bmatrix}$ $w_2 = \begin{bmatrix} -0.3 \\ 0.1 \\ 0.5 \\ 0.2 \end{bmatrix}$ $w_3 = \begin{bmatrix} 0.2 \\ -0.4 \\ 0.1 \\ 0.3 \end{bmatrix}$

then
$$\begin{cases} z_1 = w_1^T x = 1.71 \\ z_2 = w_2^T x = -0.44 \\ z_3 = w_3^T x = -0.18 \end{cases} \qquad \begin{cases} p_1 = 0.784 \\ p_2 = 0.091 \\ p_3 = 0.12 \end{cases}$$

let $y = [1, 0, 0]$ then $p(y \mid x, W) = p_1^{1} \, p_2^{0} \, p_3^{0}$

$$= 0.789$$

multiclass classification.

- $$\boxed{p_{ik} = P(y_i = k \mid x_i, W) = \frac{e^{w_k^T x_i}}{\sum_j e^{w_j^T x_i}}}$$ $$W = \begin{bmatrix} - w_1^T - \\ \vdots \end{bmatrix}$$

- $$\mathcal{L} = P(y_1, \dots, y_n \mid x_1, \dots, x_n, W)$$

$$= \prod_{i=1}^{N} p(y_i \mid x_i, W)$$

$$y_{ik} = \begin{cases} 1 & y_i = k \\ 0 & y_i \neq k \end{cases}$$

$$= \prod_{i=1}^{N} \prod_{k=1}^{C} \underbrace{p(y_i = k \mid x_i, W)}_{=: p_{ik}}{}^{y_{ik}}$$

- $$-\log \mathcal{L} = -\sum_{i=1}^{N} \sum_{k=1}^{C} y_{ik} \log p_{ik}$$

- $$p_k = \frac{e^{w_k^T x}}{\sum e^{w_j^T x}} = \frac{e^{z_k}}{\sum e^{z_j}}$$

class

- $$\frac{\partial}{\partial w_{mn}} p_k = \frac{\underbrace{\boxed{\frac{\partial e^{z_k}}{\partial w_{mn}}}}_{= \delta_{mk} \cdot e^{z_k} \cdot x_n} \cdot (\sum e^{z_j}) + (e^{z_k}) \cdot \underbrace{\boxed{\frac{\partial \sum e^{z_j}}{\partial w_{mn}}}}_{= e^{z_m} \cdot x_n}}{(\sum e^{z_j})^2}$$

class

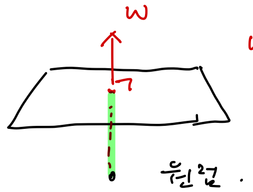
$$= p_k (\delta_{mk} - p_m) \cdot x_n$$

- $$\frac{\partial}{\partial w_{mn}}(-\log \mathcal{L}) = -\sum_{i}^{N} \sum_{k}^{C} y_{ik} \cdot \frac{1}{\cancel{p_{ik}}} \cdot \cancel{p_{ik}} (\delta_{mk} - p_{im}) x_{in}$$

$$= \sum_{i}^{N} (p_{im} - y_{im}) x_{in}$$

예측 정답

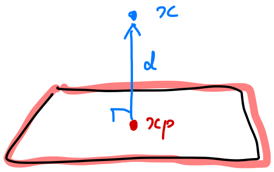
2.2



$w^Tx+b=0$

$$w^Tx+b \begin{cases} >0 & y=1 & \text{Positive} \\ <0 & y=-1 & \text{Negative} \end{cases}$$

hence $y(w^Tx+b) \geq 0$



hyperplane

$$0 = w^Tx_p + b$$
$$= w^T(x-d)+b$$
$$= w^T(x-\alpha w)+b$$

- $\|d\| = \alpha \cdot \|w\|$

$$= (w^Tx+b)/\|w\|$$

1.

$$\underset{w,b}{\arg\max} \left[\min_{x} \frac{\cancel{|w^Tx+b|}}{\|w\|} \right] = \underset{w,b}{\arg\min} \; \|w\|$$

- margin

: $r(w,b) = \min_{x\in D} \dfrac{|w^Tx+b|}{\|w\|}$

"margin을 최대화 하는 것이 SVM의 목표이다."

s.t ① $\forall i \quad y_i(w^Tx_i+b) \geq 0$

② $\min_{x} |w^Tx+b| = 1$

$\Longleftrightarrow \forall i \quad y_i(w^Tx_i+b) \geq 1$

---

minimize $\quad \frac{1}{2}\|w\|^2 + \max_{\alpha_i \geq 0} \alpha_i \left(1 - y_i(w^Tx_i+b)\right)$

$$\max_{\alpha\geq 0} \alpha\left(1-y(w^Tx+b)\right) = \begin{cases} 0 & y(w^Tx+b) \geq 1 \\ \infty & \text{else.} \end{cases}$$

- $$\min_{w,b} \max_{a_i \geq 0} \left[\frac{1}{2}\|w\|^2 + \sum_i \alpha_i \left(1 - y_i(w^T x_i + b)\right) \right]$$

$$\geq \max_{a_i \geq 0} \min_{w,b} \left[\frac{1}{2}\|w\|^2 + \sum_i \alpha_i \left(1 - y_i(w^T x_i + b)\right) \right]$$

- $$\mathcal{L} = \frac{1}{2} w^T w + \sum_i \alpha_i - w^T \left(\sum_i \alpha_i y_i x_i \right) - \sum_i \alpha_i y_i b$$

$$\begin{cases} \nabla_w \mathcal{L} = 0 & w = \sum_i \alpha_i y_i x_i \\ \dfrac{\partial \mathcal{L}}{\partial b} = 0 & \sum_i \alpha_i y_i = 0 \end{cases}$$

- $$\boxed{\max_{\alpha_i} \sum_i \alpha_i - \frac{1}{2} \overbrace{\sum_i \sum_j \alpha_i \alpha_j y_i y_j x_i^T x_j}^{w^T w}}$$

s.t. ① $\alpha_i \geq 0$

② $\sum \alpha_i y_i = 0$

- support vectors I.

① $\alpha_i > 0$ then $i \in I$

② obtain $w = \sum_{i \in I} \alpha_i y_i x_i$

③ obtain $b = y_i - x_i^T w$ $(i \in I)$

④ $$\|w\|^2 = \overbrace{\left[\sum_{i \in I} \alpha_i y_i x_i^T \right]}^{w^T} w$$

$1 = y_i(x_i^T w + b)$

$$= \sum_{i \in I} \alpha_i (1 - y_i b)$$

$$= \sum_{i \in I} \alpha_i - b \cancel{\sum_{i \in I} \alpha_i y_i}$$

$$= \sum_{i \in I} \alpha_i$$

⑤ $$\text{margin} = \frac{|x^T w + b|}{\|w\|} = \frac{1}{\sqrt{\sum_{i \in I} \alpha_i}}$$