PCR

$$V^T = \begin{bmatrix} v_1^T \\ \hline v_2^T \\ \hline v_3^T \end{bmatrix}$$

① $\underset{N\times p}{X} = \underset{N\times p}{U} \, \underset{p\times p}{D} \, \underset{p\times p}{V^T}$

$X v_1 = d_1 U_1$ (fact)

$$\begin{bmatrix} \underbrace{d_1 U_1}_{} & d_2 U_2 & \cdots \end{bmatrix}_{N\times p}$$

② $\underbrace{\qquad\qquad\qquad}_{\text{3 components}}$

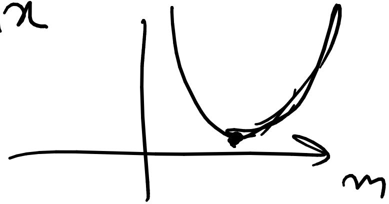
LASSO

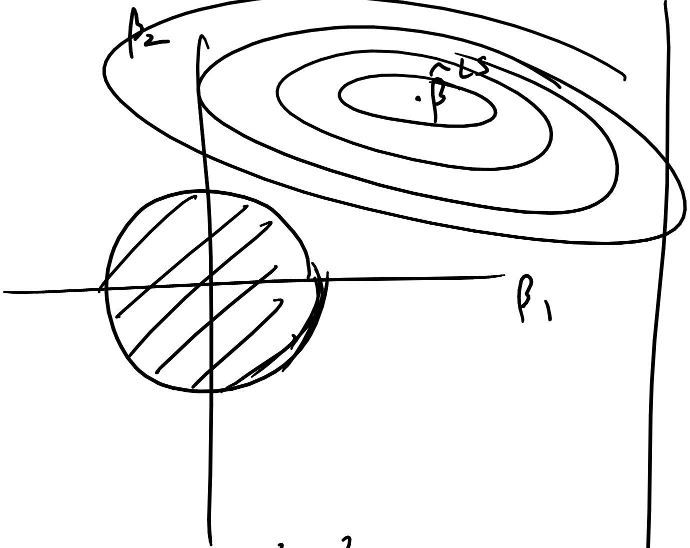
$$(Y - \bar{X}\beta)^T (Y - \bar{X}\beta) + \lambda \underbrace{\|\beta\|_1}_{= \sum_{j=1}^{p} |\beta_j|}$$

l_1 norm

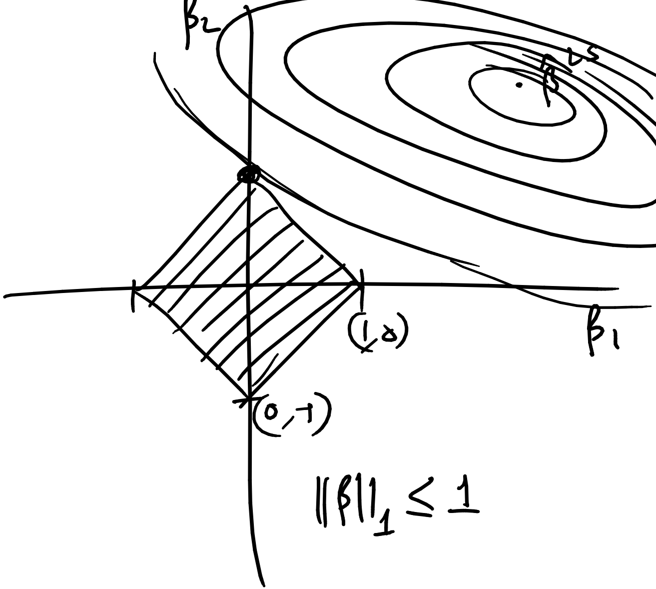
eg: $y = mx$

$$\sum_{i=1}^{n} (y_i - m x_i)^2$$





$\beta^T\beta \leq \tau \qquad \beta_1^2 + \beta_2^2 \leq 1$



$\|\beta\|_1 \leq 1$

1. $X, G \left\{ \begin{array}{l} 1 \\ 2 \\ \vdots \\ K \end{array} \right.$

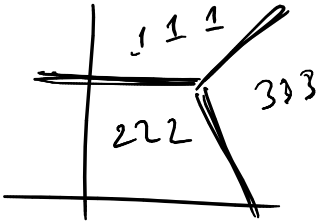
$$\begin{array}{ccccc} Y_1 & \cdots & & & Y_K \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ & & & & \end{bmatrix}$$

G data

$$\begin{bmatrix} 1 \\ 2 \\ 1 \\ 3 \\ 2 \\ \vdots \end{bmatrix}$$

$N \times 1$



$\hat{\beta}_K^T X$

$\hat{\beta}_K^T X = \hat{\beta}_\ell^T X$

hyperplane

discriminant functions.

$$\underset{p \times K}{\hat{\beta}} = (X^T X)^{-1} X^T \underset{N \times K}{Y}$$

issue 1 : $\hat{\beta}_k^T X \in [0,1]$

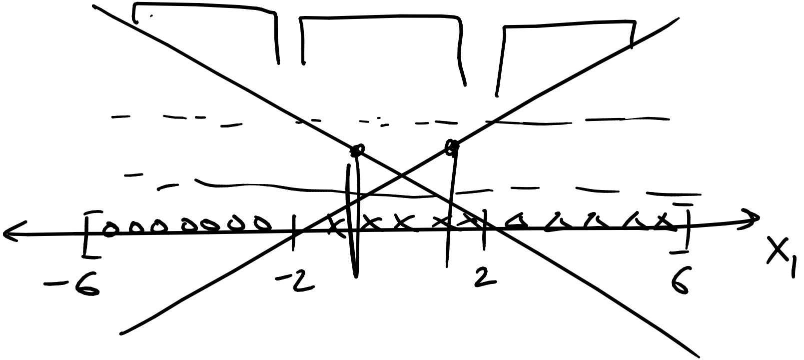
issue 2 : class masking

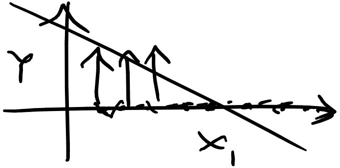
$E[Y \mid X = x] = \beta^T x$

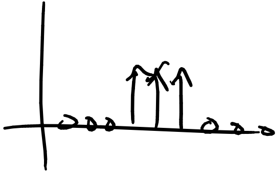
X, Y

$P(G = k \mid X = x)$

$= P(Y_k = 1 \mid X = x)$







$$\underset{k=1,..K}{\operatorname{argmax}}\ \hat{\beta}_k^T X$$

$$\underline{\underline{P_{X,G}}}$$

$$\left[\underbrace{P(G=k \mid X=x)} = \frac{\overbrace{P(X=x \mid G=k)}^{\prod_{j=1}^{p} P(X_j = x_j \mid G=k)} \; P(G=k)}{P(X=x)} \right]$$

$$\pi_k \quad (1)$$

$$P(X=x \mid G=k) = N(\mu_k, \Sigma) \quad (2)$$

$$\pi_1,\ \pi_2,\ \pi_3,\ \underset{p\times 1}{\mu_1},\ \underset{p\times 1}{\mu_2},\ \underset{p\times 1}{\mu_3},\ \underset{p\times p}{\Sigma}$$

$$\log\left(\frac{P(G=1 \mid X=x)}{P(G=2 \mid X=x)}\right) = 0$$

$$g(x) = c^T x$$

$$\boxed{\frac{1}{\sqrt{2\pi \det(\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}$$

QDA: $\pi_1, \pi_2, \pi_3, \mu_1, \mu_2, \mu_3, \Sigma_1, \Sigma_2, \Sigma_3$

$p = 2$

LDA: $X^{\text{original}}_{2\times 1} \longrightarrow \begin{bmatrix} X_1 \\ X_2 \\ X_1 X_2 \end{bmatrix}$

π_1', π_2', π_3'

$\mu_1', \mu_2', \mu_3' \in \mathbb{R}^3$

$\Sigma'_{3\times 3}$

Computation: "easy" due to MLE

Logistic Regression

3 classes

$$\log\left(\frac{P(G=1|X=x)}{P(G=3|X=x)}\right) = \beta_1^T x$$

$$\log\left(\frac{P(G=2|X=x)}{P(G=3|X=x)}\right) = \beta_2^T x$$

$$\Rightarrow \quad P(G=k|X=x) = \frac{e^{\beta_k^T x}}{1+e^{\beta_1^T x}+e^{\beta_2^T x}}$$

loss : $\underbrace{\sum_{i=1}^{N}}$ $\log \underbrace{P(G = g_i \mid X = x_i)}$ $\{x_i, g_i\}_{i=1}^{N}$

Same function of β_1, β_2 for each observation i.

in example $\log P(g_i \mid x_i) = \underline{\sum_{j=1}^{p} y_{ij} \log P(y_{ij} \mid x_i)}$

3 classes $g_i = 1$ $y_i^T = [1\ 0\ 0]$

y_{i1} y_{i2} y_{i3}

$\underline{\sum_{j=1}^{p} q_j \log q_j}$

negative of entropy

AIC, BIC

$$\underbrace{\frac{1}{N}\text{Training error}}_{\overline{\text{err}}} + \frac{2}{N}\left(g\left(\frac{1}{\lambda}\right)\right)$$

$$\frac{2\log N}{N}\left(g\left(\frac{1}{\lambda}\right)\right)$$

$$\log_e a = \frac{\log_2 a}{\log_2 e} = \log_2 2.73 = 1.6$$