problem 2 - HW3 - chap6 (ISL book) $\boxed{5}$-d)

5)
d)

$$L(\beta)=(y_1-\beta_1 x_1-\beta_2 x_1)^2+(y_2-\beta_1 x_2-\beta_2 x_2)^2 \quad s.t \quad |\beta_1|+|\beta_2|<t$$

$$L(\beta)=(y_1-\beta_1 x_1-\beta_2 x_1)^2+(\overset{-y_1}{y_2}-\beta_1 \overset{-x_1}{x_2}-\beta_2 \overset{-x_1}{x_2})^2 \xrightarrow[x_2=-x_1]{y_2=-y_1}$$

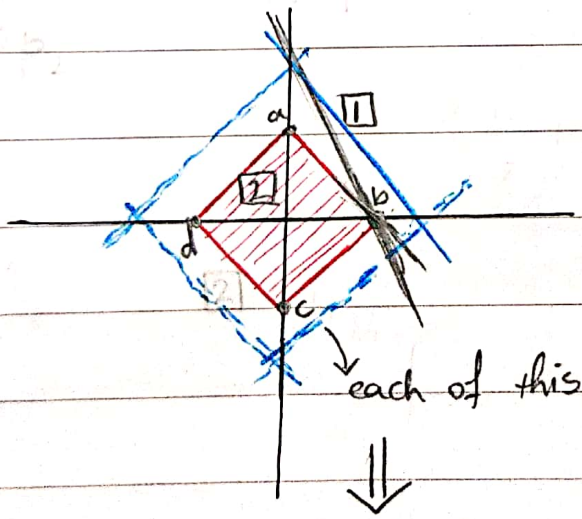
$$\Rightarrow L(\beta)=(y_1-\beta_1 x_1-\beta_2 x_1)^2+(-y_1+\beta_1 x_1+\beta_2 x_1)^2$$

$$L(\beta)=2y_1^2+2\hat{\beta}_1^2 x_1^2+2\hat{\beta}_2^2 x_1^2-4\beta_1 x_1 y_1-4\beta_2 x_1 y_1+4\beta_1\beta_2 x_1^2$$

$$\frac{\partial L(\beta)}{\partial \beta_1}=0 \rightarrow 4\hat{\beta}_1 x_1^2-4x_1y_1+4\beta_2 x_1^2=0 \rightarrow 4x_1^2(\hat{\beta}_1+\hat{\beta}_2)=4x_1y_1$$

$$\Downarrow$$

$$\boxed{1}\quad \boxed{\hat{\beta}_1+\hat{\beta}_2=\frac{y_1}{x_1}}$$



constraint

$$\boxed{2}\quad |\beta_1|+|\beta_2|<t$$

each of this lines can show the $\boxed{1}$ formula.

$$\Downarrow$$

we should change β_1 & β_2 until line 1 crosses one of the points (a-d)

$$\Downarrow$$

the intersection point is the lasso minimization for β parameters.

Hw3 / ISL-book

~~problem 2~~ chap6 $\boxed{7}$

a, b, e)

guasian dist.

$$\boxed{p(x) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{1}{2\delta^2}(x-\mu)^2\right)}$$

likelihood $\rightarrow p(y[1], \dots, y[n] \mid x, \beta)$

$$\text{Data} = \{y[1], \dots, y[n]\} \rightarrow \overset{\text{likelihood}}{L(\mu)} = \prod p(y[i])$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\delta}\right)^n \exp\left(-\frac{1}{2\delta^2}\sum\left(y[i] - \beta^T x[i]\right)^2\right)$$

prior (according to guasian dist) $\propto \exp\left(\frac{-1}{2\delta^2}\sum(\beta_i - \hat{\mu})^2\right)$

$\beta_i \sim N(0, T^2)$

$$= \exp\left(-\frac{1}{2\delta^2}\sum \beta_i^2\right)$$

posterior $\propto$ prior $\times$ likelihood $\propto$ log prior + log likelihood

$$= \frac{-1}{2\delta^2}\sum\left(y[i] - \beta^T x[i]\right)^2 - \frac{1}{2T^2}\sum \beta_i^2$$

$$= \sum\left(y[i] - \beta^T x[i]^2\right) + \boxed{\frac{\delta^2}{T^2}}\sum \beta_i^2$$

اگر این معادل λ در نظر گرفته شود، عبارت کلی همان $\hat{\beta}$ ridge است

چون در اینجا یک posterior gausian dist محسوب می‌شود، پس در اینجا mode و میانه

برای β همان تخمین ridge regression است. چون در توزیع نرمال $\leftarrow$ $\boxed{\text{mode} = \text{mean}}$

problem 2 - chap6 - 7

c) - d)

likelihood $\propto \exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(y[i]-\beta^T x[i])^2\right)$

prior → laplace (double exponential) distribution :

$$P(\beta)=\prod P(\beta_i) \propto \exp\left(\frac{-1}{b}\sum|\beta_i-\mu|\right)$$

[because laplace dist is : $\propto \boxed{\frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right)}$]

posterior = prior x likelihood

= log prior + log likelihood

$$\propto \frac{-1}{2\delta^2}\sum_{i=1}^{n}(y[i]-\beta^T x[i]^2)+\frac{-1}{b}\sum|\beta_i-\mu|$$

$$\boxed{\sum_{i=1}^{n}(y[i]-\beta^T x[i]^2)+\boxed{\frac{\delta^2}{b}}\sum|\beta_i-\mu|}$$

اگر این عبارت معادل λ در نظر گرفته شود، این عبارت - همان $\hat{\beta}_{lasso}$ است.

⇓

پس در اینجا mode برای β (همان posterior)، همان - lasso minimization است. ✓

/ /

problem 5 : Maximum likelihood estimation of multinomial distribution.

$x \to$ random variable $\to x_1, x_2, \dots, x_K$

have $\underline{K}$ parameters $\to \Pi = (\pi_1, \dots, \pi_K)$, $P(x = x_K) = \pi_K$

subject to $\to \sum_K \pi_K = 1 \to \sum \pi_K - 1 = 0$

- observed data $\to (n_1, \dots, n_K) \to n_K$: the number of times the value x_K appears in the data

prove the MLE for π_k is $\frac{n_K}{n}$ where $n = \sum n_K$

$$\underbrace{P(n_1, \dots, n_K \mid \pi_1, \dots, \pi_K)}_{P(D|\Theta)} = \binom{n}{n_1, n_2, n_K} \pi_1^{n_1} \pi_1^{n_2} \dots \pi_K^{n_K} = \frac{n!}{\prod n_i!} \prod \pi_i^{n_i}$$

$$l(\overbrace{\pi_1, \dots, \pi_K}^{\Theta}) = \log n! - \sum_{i=1}^{K} \log n_i! + \sum_{i=1}^{K} n_i \log \pi_i$$

$$L(\pi_1, \dots, \pi_K, \lambda) = l(\pi_1, \dots, \pi_K) + \lambda \left(\sum_{i=1}^{K} \pi_i - 1 \right)$$

lagrangian

$$= \log n! - \sum_{i=1}^{K} \log n_i! + \sum_{i=1}^{K} n_i \log \pi_i + \lambda \left(1 - \sum_{i=1}^{K} \pi_i \right)$$

$$\frac{\partial L(\pi_1, \dots, \pi_K, \lambda)}{\partial \pi_i} = 0 \to \sum_{i=1}^{K} n_i \frac{1}{\pi_i \ln 10} + \lambda = 0 \Rightarrow \lambda = \frac{n_i}{\pi_i} \to \pi_i = \frac{n_i}{\lambda}$$

از تک تک π_i مشتق گرفته

$$\frac{\partial L(\pi_1, \dots, \pi_K, \lambda)}{\partial \lambda} = 0 \to \sum_{i=1}^{K} \pi_i - 1 = 0 \to \sum \pi_i = 1$$

$$\sum \pi_i = \sum \frac{n_i}{\lambda} \to \boxed{\lambda = n} \to \boxed{\pi_i = \frac{n_i}{n}}$$