# #1-1

1) $L(W) = \Sigma(y_i - x_i^T W)^2 = \|y - XW\|^2$

$\nabla_W L(W) = -2X^T(y - XW)$

$= +2X^TXW - 2X^Ty$

$\therefore \hat{W}_{OLS} = (X^TX)^{-1}X^Ty$

2) $p(y_i | x_i ; W) = \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\dfrac{(y_i - x_i^T W)^2}{2\sigma^2}\right)$

$\log L(W, \sigma^2) = -\dfrac{n}{2}\log(2\pi\sigma^2) - \dfrac{1}{2\sigma^2}\Sigma(y_i - x_i^T W)^2$

$\therefore \hat{W}_{MLE} = \hat{W}_{OLS} = (X^TX)^{-1}X^Ty$

3) OLS와 MLE 모두

$\Sigma(y_i - x_i^T W)^2 = \|y - XW\|^2$

$= y^Ty - 2y^TXW + W^TX^TXW$

$\nabla^2 f(W) = 2X^TX \succeq 0$

∴ 모두 convex function $\Sigma(y_i - x_i^T W)^2$을 최소화하는 문제로 변환.

$\therefore \hat{W}_{OLS} = \hat{W}_{MLE}$

# # 1-3.

1) Ridge: 경계가 매끈. 해가 축에 딱 붙기 어려움.

⇒ 모든 계수를 조금씩 줄여 분산↓. 예측안정성↑

LASSO: 꼭짓점이 축 위에 있음 → 해가 꼭짓점에서. 0이 되게됨.

⇒ 쓸모없는 계수는 0으로 만들어 차원 자체를 줄임.

2) Ridge 목적함수 : $\|y - X\beta\|^2 + \lambda\Sigma\beta_j^2$ : 전부 매끈한 미분가능 이차함수

$\nabla = 0 \rightarrow \beta = (X^TX + \lambda I)^{-1}X^Ty$

LASSO 목적함수 : $\|y - X\beta\|^2 + \lambda\Sigma|\beta_j|$

$|\cdot|$ 0에서 꺾임 → 부호를 나눠줘야함.

## #2-1.

$$\min_{W,b,\{\varepsilon_i\}} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N}\varepsilon_i . \quad y_i(W^T x_i + b) \geq 1-\varepsilon_i . \ \varepsilon_i \geq 0 . \ \forall i = 1 \dots N$$

W.b 고정. 각 i에 대해 $\varepsilon_i$만 최소화

$$\rightarrow \min_{\varepsilon_i} C\varepsilon_i . \quad \varepsilon_i \geq 0 . \ \varepsilon_i \geq 1 - y_i(W^T x_i + b)$$

$$\min_{W,b} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{N} \max(0, 1-y_i(W^T x_i + b))$$

## #2-2.

$$L(W,b) = \frac{1}{2}\|W\|^2 + C \cdot \max(0, 1 - y_i(W^T x_i - b))$$

$$y_i(W^T x_i - b) \geq 1 : \ L(W,b) = \frac{1}{2}\|W\|^2$$

$$y_i(W^T x_i - b) < 1 : \ L(W,b) = \frac{1}{2}\|W\|^2 + C(1 - y_i(W^T x_i - b))$$

$$\therefore \frac{dL}{dW} = \begin{cases} W & \text{if } \ y_i(W^T x_i - b) \geq 1 \\ W - C \cdot y_i x_i & \text{if } \ y_i(W^T x_i - b) < 1 \end{cases}$$

$$\therefore \frac{dL}{db} = \begin{cases} 0 & \text{if } \ y_i(W^T x_i - b) \geq 1 \\ C \cdot y_i & \text{if } \ y_i(W^T x_i - b) < 1 \end{cases}$$

#3-1.

$$\begin{cases} \log p(X|\theta) = E_q[\log p(X,Z|\theta)] - E_q[\log p(Z|X,\theta)] \\ q(Z) \equiv p(Z|X,\theta^{(t)}) \end{cases}$$

$$\rightarrow \log p(X|\theta) = E_{p(Z|X,\theta^{(t)})}[\log p(X,Z|\theta)] - E_{p(Z|X,\theta^{(t)})}[\log p(Z|X,\theta)]$$

$$Q(\theta|\theta^{(t)}) \equiv E_{p(Z|X,\theta^{(t)})}[\log P(X,Z|\theta)]$$

$$\therefore \log p(X|\theta) = Q(\theta|\theta^{(t)}) - E_{p(Z|X,\theta^{(t)})}[\log p(Z|X,\theta)]$$

$$\log p(X|\theta) = E_q[\log p(X,Z|\theta)] - E_q[\log q(Z)] + \underbrace{E_q[\log q(Z)] - E_q[\log p(Z|X,\theta)]}_{KL(q\|p)}$$

$$\log p(X|\theta) \geq E_q[\log p(X,Z|\theta)] - E_q[\log q(Z)]$$

$$= Q(\theta|\theta^{(t)}) - E_q[\log q(Z)]$$

$$\therefore E_q[\log p(X,Z|\theta)] - E_q[\log q(Z)] \uparrow$$

$$= Q(\theta|\theta^{(t)}) \uparrow$$

$$\therefore \text{M-step}: \theta^{(t+1)} = \arg\max_\theta Q(\theta|\theta^{(t)})$$

#3-2.

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{k,i,\theta^{(t)}} \log[w_k N(x_i|\mu_k, \Sigma_k)]$$

$w_k$만 포함하는 항

$$Q_w(w) = \sum_{i=1}^{n}\sum_{k=1}^{K} \gamma_{k,i,\theta^{(t)}} \log w_k$$

$$N_k^{(t)} \equiv \sum_{i=1}^{n} \gamma_{k,i,\theta^{(t)}}$$

$$\therefore Q_w(w) = \sum_{k=1}^{K} N_k^{(t)} \log w_k$$

$$\Sigma_k w_k = 1, \ w_k \geq 0$$

$$L(w,\lambda) = \sum_{k=1}^{K} N_k^{(t)} \log w_k + \lambda\left(\sum_{i=1}^{K} w_k - 1\right)$$

$$\frac{\partial L}{\partial w_k} = \frac{N_k^{(t)}}{w_k} + \lambda = 0 \rightarrow w_k = -\frac{N_k^{(t)}}{\lambda} \qquad \therefore w_k^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_{k,i,\theta^{(t)}}$$

$$\sum_{k=1}^{K} w_k = -\frac{1}{\lambda}\sum_{k=1}^{K} N_k^{(t)} = -\frac{1}{\lambda}\sum_{i=1}^{n}\underbrace{\sum_{k=1}^{K} \gamma_{k,i,\theta^{(t)}}}_{=1} = -\frac{n}{\lambda} = 1 \rightarrow \lambda = -n$$