행벡터 1행×n형태 $[\ \_\_\_\_\ ]$

열벡터 n행×1열형태 $\begin{bmatrix} - \\ - \\ - \end{bmatrix}$

미분한 값은 열벡터 형태로 써야함!

1-1. (1)

$$L(w) = (y - Xw)^T (y - Xw)$$

$\because (Xw)^T = w^T X^T$

$$= y^T y - \underbrace{y^T X w}_{①} - w^T X^T y + w^T X^T X w$$

둘다 스칼라 (①: $1\times n \cdot n\times d \cdot d\times 1 = 1\times 1$) (②: $1\times d \quad d\times n \quad n\times 1 = 1\times 1$) ∴ 같음 ⇒ $2y^T X w$

$$\frac{dL(w)}{w} = 0 - 2y^T X + 2X^T X w$$

$1\times n \quad n\times d$ $d\times n \quad n\times d \cdot d\times 1$ 열벡터로 써야함 (n행×1열)

$$= 0 - 2X^T y + 2X^T X w = 0$$ 인 w를 찾으면 됨

$$\cancel{2}X^T X w = \cancel{2}X^T y$$

$$\boxed{\hat{w} = (X^T X)^{-1} X^T y}$$

1-1 (2)

$$L(w) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i^T w)^2}{2\sigma^2}\right)$$

$$\log L(w) = \underline{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i^T w)^2} + \log \frac{1}{\sqrt{2\pi\sigma^2}}$$

MLE(최대우도추정) 추정값 ⇒ 로그우도가 최대가 되는 w값 찾는 문제

⇒ ___ 밑줄이 최소가 되는 w를 찾는 문제

$\sum_{i=1}^{n} (y_i - x_i^T w)^2 = (y - Xw)^T (y - Xw)$ 이 최소가 되는 w값

$$\boxed{\therefore \hat{w} = (X^T X)^{-1} X^T y}$$

1-1 (3)

OLS와 MLE 모두 convex function을 최소화 하는 문제로 바뀌고 그 결과값은 $(X^T X)^{-1} X^T y$로 서로 같게 된다.

1-3 e

1) Ridge L_2 패널티 vs Lasso L_1 패널티

$\lambda \sum_{j=1}^{p} B_j^2$ 형태 vs $\lambda \sum_{j=1}^{p} |B_j|$ 형태

L_2 제약은 기하학적으로 원형태이므로, 모서리가 없어 계수들이 연속적으로 줄어든다

L_1 제약은 마름모 형태이므로, 해가 모서리에 닿게 되면 계수가 정확히 0이 될 수 있다

즉 Ridge는 계수를 0으로 만들지는 못하지만 다중공선성을 완화할 수 있고

Lasso는 불필요한 변수를 0으로 만들 수 있어 변수선택에 용이하다

2) Ridge의 L2 패널티는 B_j^2 형태로 미분하여 0이 되는 해가 존재한다. 따라서 해석해가 존재할 수 있다.

반면에 Lasso의 L_1 형태는 $|B_j|$ (절대값) 형태로 ~~미분할 수 없는 지점이 존재하기~~ 미분하여 0이 되는 곳이 없기에 해석해가 없다

2-1

$\varepsilon_i \geq 0$ 이고, $\varepsilon_i \geq 1 - y^{(i)}(w^T x^{(i)} + b)$ 이므로

$\min\left(C \sum_{i=1}^{n} \varepsilon_i\right) = \min\left(C \sum_{i=1}^{n} \max(0, 1 - y^{(i)}(w^T x^{(i)} + b))\right)$ 이다

$\Rightarrow$ ε_i를 가질 수 있는 가장 작은 값은 0과 $1 - y^{(i)}(w^T x^{(i)} + b)$ 중 큰 값이다

2-2

$L(w, b) = \underline{\frac{1}{2}||w||^2} + C \cdot \underline{\max(0, 1 - y_i(w^T x_i - b))}$ → Z 라고 하자

마진을 크게 만드는데 $\quad$ $1 - y_i(w^T x_i - b) > 0$ 에선 벌점을 부과함

1) $\frac{1}{2}||w||^2$ 미분 $= \frac{d}{dw} \frac{1}{2} w^T w = w$

2) $1 - y_i(w^T x_i - b)$ 미분 (w에 대해) $\Rightarrow \frac{d}{dw}(1 - y_i(w^T x_i - b)) = -y_i x_i$

$$\therefore \frac{dL}{dw} = \begin{cases} w & (Z = 0 \;\because\; 1 - y_i(w^T x_i - b) \leq 0 = y_i(w^T x_i - b) \geq 1) \\ w + C \times (-y_i x_i) & (Z = \underbrace{1 - y_i(w^T x_i - b)}_{>0} \text{ 즉 } y_i(w^T x_i - b) < 1 \end{cases}$$

$$\frac{dL}{db} = \begin{cases} 0 & \text{if } 1 - y_i(w^T x_i - b) \leq 1 \\ C \frac{d}{db}(y_i b) = C \times y_i & \text{if } 1 - y_i(w^T x_i - b) > 0 \end{cases}$$

3-1

$$\log P(X|\theta) = \log \sum_z q(z) \frac{P(X,z|\theta)}{q(z)} \overset{\text{Jensen}}{\geq} \underline{\sum_z q(z) \log \frac{P(X,z|\theta)}{q(z)}}$$

$$\underbrace{\sum_z q(z) \log \frac{p(X,z|\theta)}{q(z)}}_{F(q,\theta)} = \sum_z q(z) \log p(X,z|\theta) - \sum_z q(z) \log q(z)$$

$$\sum_z q(z) \log p(X,z|\theta) = \sum_z q(z) \left[\log p(z|X,\theta) + \log p(X|\theta) \right]$$

$$= \sum_z q(z) \log p(z|X,\theta) + \sum_z q(z) \log p(X|\theta)$$

$$= \sum_z q(z) \log p(z|X,\theta) + \log p(X|\theta)$$

$$F(q,\theta) = \sum_z q(z) \log p(z|X,\theta) + \log p(X|\theta) - \sum_z q(z) \log q(z)$$

$$= \log p(X|\theta) + \sum_z q(z) \log p(z|X,\theta) - \sum_z q(z) \log q(z)$$

$$= \log p(X|\theta) - KL(q(z) \| p(z|X,\theta))$$

$$\log p(X|\theta) = F(q,\theta) + \underbrace{KL(q(z) \| p(z|X,\theta))}_{\geq 0}$$

$KL(q(z) \| p(z|x,\theta)) = 0 \iff q(z) = p(z|x,\theta)$

$\therefore$ E-step에서 $q(z) = p(z|x,\theta^{(t)})$ 일때

$\log p(x|\theta) = F(q,\theta)$ 이 된다 (최소가 된다)

이때의 $q(z) = p(z|x,\theta)$ 를 E-step에서 택한다