Exercise. Chap 1. 정성규 교수님 응용통계 9장 참고.

1.2.

for a pre-specified $\lambda > 0$

$$\hat{\beta}^R = \underset{\beta}{\arg\min} \left\{ (y_c - X_c\beta)^T (y_c - X_c\beta) + \lambda \beta^T \beta \right\}.$$

(where. $y = 1 \cdot \alpha_0 + X_c\beta + \varepsilon$, $y - \bar{y} = y_c = X_c\beta + \varepsilon$).

$$= \underset{\beta}{\arg\min} \{ \underbrace{SSE(\beta)}_{Loss} + \underbrace{\lambda \|\beta\|^2}_{penalty} \}.$$

최소화할 식을 미분하면

$$-2\beta^T X_c^T y_c + 2X_c^T X_c \beta + 2\lambda\beta = 0.$$

$$\therefore \hat{\beta}^R = (X_c^T X_c + \lambda I)^{-1} X_c^T y.$$

* woodbury formula.

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$$

$$\hat{\beta}^R = \left\{ (X_c^T X_c)^{-1} - (X_c^T X_c)^{-1} \left(\frac{1}{\lambda} I + (X_c^T X_c)^{-1} \right)^{-1} (X_c^T X_c)^{-1} \right\} X_c^T y$$

$$= \left(I - (X_c^T X_c)^{-1} \left(\frac{1}{\lambda} I + (X_c^T X_c)^{-1} \right)^{-1} \right) \hat{\beta}$$

$$= \left\{ I - \lambda (X_c^T X_c)^{-1} \left(I + \lambda (X_c^T X_c)^{-1} \right)^{-1} \right\} \hat{\beta}$$

$$= \left\{ I - \left(\lambda (X_c^T X_c)^{-1} + I - I \right) \left(I + \lambda (X_c^T X_c)^{-1} \right)^{-1} \right\} \hat{\beta}$$

$$= \left\{ I - I + \left(I + \lambda (X_c^T X_c)^{-1} \right)^{-1} \right\} \hat{\beta}$$

$$= \left(I + \lambda (X_c^T X_c)^{-1} \right)^{-1} \hat{\beta}$$

1.8.

$$\int x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx \qquad \frac{x-\mu}{\sigma} := Y$$

$$= \int (\sigma Y+\mu) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}Y^2\right) \cdot \sigma \, dY$$

$$= \int (\sigma Y+\mu) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Y^2\right) dY$$

$$= \int \sigma \cdot Y \cdot \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Y^2\right) dY + \mu \int \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}Y^2\right) dY$$

$$= \mu.$$

$$\int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx = \int A(\sigma^2)\, B(\sigma^2)\, dx = 1.$$

$$\frac{d}{d\sigma^2} \int A(\sigma^2) B(\sigma^2)\, dx = 0.$$

$$\int \frac{d}{d\sigma^2} \left(A(\sigma^2) B(\sigma^2)\right) dx = 0 \qquad (\sigma^2\text{으로 미분!!})$$

$$\int A'(\sigma^2) B(\sigma^2) + A(\sigma^2) B'(\sigma^2)\, dx = 0.$$

$$\int \frac{1}{\sqrt{2\pi}} \left(-\frac{1}{2}\right) \left(\frac{1}{\sigma^2}\right)^{\frac{3}{2}} \cdot \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) + \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \left(\frac{x-\mu}{\sigma}\right)^2 \cdot \frac{1}{2\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \left(-\frac{1}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\sigma^4}(x-\mu)^2 \right) dx = 0.$$

$$\Rightarrow \quad -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} E(X^2) - \frac{2\mu}{2\sigma^4} E(X) + \frac{\mu^2}{2\sigma^4} = 0.$$

$$\therefore E(X^2) = \mu^2 + \sigma^2.$$

1.10

$X \perp Z$.

① $E(X+Z) = E(X) + E(Z)$ (적분으로 쉽게 보일수 O.)

② $Var(X+Z) = Var(X) + Var(Z) + 2Cov(X, Z)$ (이것도 적분으로 보일수 O)

1.15

중복조합.

변수 $X = (x_1, x_2, \cdots x_D)$ 의 원소 조합을 선택하되 차수의 합이 M이

되도록 만들기.

즉 $\sum^D \sum^D \cdots \sum^D w_{i_1 i_2 \cdots i_M} x_{i_1} x_{i_2} \cdots x_{i_M} \Rightarrow \sum_{i_1=1}^{D} \sum_{i_2=1}^{i_1} \cdots \sum_{i_M=1}^{i_{M-1}} \tilde{w}_{i_1 \cdots i_M} x_{i_1} x_{i_2} \cdots x_{i_M}$

이 때 가능한 x 들의 조합은 중복조합으로 추출가능.

D개 중에 M개를 중복 허용해서 추출 : ${}_D H_M = {}_{(D+M-1)} C_M$

$= {}_{(D+M-1)} C_{(D-1)}$.

공식 유도 배경 : D개 中 M개를 중복허용 추출

⟺ M 개가 1열로 나열되어 있고 $(D-1)$개의 칸막이를 설치하는 것.

ex) // o // o o o o o o o

$\underbrace{M = 7, \quad D = 5}$

⟺ ${}_{7+5-1} C_{5-1}$

또한 책에서는 ${}_D H_M = \sum_{i=1}^{D} {}_i H_{M-1}$ 로 제시하는데

예를들어 $D=3$, $M=3$ 인 경우.

$D=1$, $M=2$ $\quad x_1 x_1$

$D=2$, $M=2$ $\quad x_1 x_1$, $x_1 x_2$, $x_2 x_2$

$D=3$, $M=2$ $\quad x_1 x_1$, $x_1 x_2$, $x_1 x_3$, $x_2 x_2$, $x_2 x_3$, $x_3 x_3$

⇒ $x_1 x_1$은 공통인데 이때는 해당원소보다 큰 x_1, x_2, x_3를 넣어주면 각기 다른것. $x_1 x_2$는 $x_1 x_2 x_2$, $x_1 x_2 x_3$. 즉 가장 큰 원소 이상의 것을 붙여준다.

1.21.

(i) $a, b \geq 0$,

$$a \leqq b \Rightarrow \cancel{a = \frac{a}{2} + \frac{a}{2} \leqq \frac{a}{2} + \frac{b}{2} \leqq \sqrt{ab}} \text{ (~~산술·기하 평균~~)}$$

$$\sqrt{a} \leqq \sqrt{b} \Rightarrow a = \sqrt{a}\sqrt{a} \leqq \sqrt{a}\sqrt{b} = (ab)^{\frac{1}{2}}$$

(ii)
$$p(\text{mistake}) = \int_{R_2} p(X, C_1)\, dx + \int_{R_1} p(X, C_2)\, dx$$

$$\leqq \int_{R_2} p(X, C_1)^{\frac{1}{2}} p(X, C_2)^{\frac{1}{2}}\, dx + \int_{R_1} p(X, C_2)^{\frac{1}{2}} p(X, C_1)^{\frac{1}{2}}\, dx$$

$$= \int \{p(X, C_1)\, p(X, C_2)\}^{\frac{1}{2}}\, dx.$$

1.25

$$E[L(t, y(x))] = \iint \| y(x) - t \|^2 p(x, t)\, dx\, dt$$

$$= \iint (y(x) - t)^2 p(x, t)\, dx\, dt.$$

$$\frac{d}{dy(x)} E[L(t, y(x))] = \iint 2(y(x) - t)\, p(x, t)\, dx\, dt = 0.$$

즉 $\iint y(x)\, p(x, t)\, dx\, dt = \iint t\, p(x, t)\, dx\, dt$

$$\iint y(x)\, p(x)\, \underbrace{p(t|x)}_{\text{1로}}\, dx\, dt = \iint t\, p(x)\, p(t|x)\, dx\, dt.$$

$$y(x) = \int p(x)\, dx \int t\, p(t|x)\, dt = E(t|x).$$

※ 풀이에서 처럼 $\int 2(y(x) - t)\, p(x, t)\, dt = 0$. 을 구하면 됨. (변수 t로 적분 되는 부분 신경쓰기)

1.33.

$$H[y|x] = -\sum\sum p(x,y)\ln p(x,y) - \left(-\sum p(x)\ln p(x)\right)$$

①

$$= -\sum\sum p(y|x)\ln p(y|x) = 0.$$

1)

① 이 가능한 이유

$$-\sum_Y\sum_X p(Y|X)p(X)\ln p(Y|X)p(X) - \left(-\sum p(X)\ln p(X)\right)$$

$$= -\sum_Y\sum_X p(Y|X)p(X)\left\{\ln p(Y|X) + \ln p(X)\right\} + \sum p(X)\ln p(X)$$

$$= -\sum_Y p(Y|X)\ln p(Y|X)$$

2) 또한 $\sum -z_i\ln z_i = 0$ $(0\le z_i\le 1)$ 이라면 필연적으로

$z_i = 0.$ or $z_i = 1.$

따라서, $p(Y|X) = 0$ or $1.$

그런데 $p(Y|X)$는 확률이므로 하나의 Y에 대해 $p(Y|X) = 1$.

⇒ Conditional entropy = 0 이면 y는 x의 함수이다.

함수 관계이므로 추가적인 정보가 없다.