Theoretical

Name: Xun Xue
UNI: xx2241

(i) $A_{LS} = \arg\min_A L_{LS}$

$$\frac{dL_{LS}}{dA} = -2\sum_{i=1}^{m}(y^i - Ax^i)x^{iT} = 0$$

$$\sum_{i=1}^{m} y^i x^{iT} = A\sum_{i=1}^{m} x^i x^{iT}$$

$\because Y = [y^1\ y^2 \dots y^m] \quad X = [x^1, x^2, \dots, x^m]$

$\therefore Y \cdot X^T = A \cdot X \cdot X^T$

$\therefore A_{LS} = (Y \cdot X^T)(X \cdot X^T)^{-1}$

(ii)
$$\frac{dL_r}{dA} = \frac{d\left(\lambda Tr(A^TA) + \sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)\right)}{dA}$$

$$= 2\lambda A - 2\sum_{i=1}^{m}(y^i - Ax^i)x^{iT}$$

$$= 2\lambda A - 2\sum_{i=1}^{m} y^i x^{iT} + 2A\sum_{i=1}^{m} x^i x^{iT} = 0$$

$$\therefore \left(2\lambda I + 2\sum_{i=1}^{m} x^i x^{iT}\right)A = 2\sum_{i=1}^{m} y^i x^{iT}$$

$\because Y = [y^1\ y^2 \dots y^m] \quad X = [x^1\ x^2 \dots x^m]$

$\therefore A_r = (Y \cdot X^T)(\lambda I + X \cdot X^T)^{-1}$

Name: Xun Xue
UNI: xx2241

(iii) $P(X,Y|A) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\sigma^2 I|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\sum_{i=1}^{m}(y^i - Ax^i)^T(\sigma^2 I)^{-1}(y^i - Ax^i)\right)$

$$\frac{d \log P(X,Y|A)}{dA} = \frac{d\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y^i - Ax^i)^T I^{-1}(y^i - Ax^i)\right)}{dA}$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{m}(y^i - Ax^i)x^{iT} = 0$$

$\therefore \sum_{i=1}^{m} Ax^i x^{iT} = \sum_{i=1}^{m} y^i x^{iT}$ $\quad \because Y = [y^1 y^2 \dots y^m] \quad X = [x^1 x^2 \dots x^m]$

$\therefore A_{ML} = Y \cdot X^T (X \cdot X^T)^{-1}$

(iv) $P(A|(X,Y)) = \frac{P((X,Y)|A)\,P(A)}{P(X,Y)}$

$$\frac{d \log P(A|(X,Y))}{dA} = \frac{d \log P((X,Y)|A)}{dA} + \frac{d \log P(A)}{dA}$$

$$P(A) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\mathrm{Tr}[\lambda(A-M)^T(A-M)]\right)$$

$$\therefore \frac{d \log(P(X,Y)|A)}{dA} + \frac{d \log P(A)}{dA} = \frac{1}{\sigma^2}\sum_{i=1}^{m}(y^i - Ax^i)x^{iT} - \lambda(A-M)$$

$\lambda M + \frac{1}{\sigma^2}\sum_{i=1}^{m} y^i x^{iT} = \left(\lambda I + \frac{1}{\sigma^2}\sum_{i=1}^{m} x^i x^{iT}\right)A$ $\quad \because Y = [y^1 y^2 \dots y^m]$, $X = [x^1 x^2 \dots x^m]$

$\therefore A_{MAP} = (Y \cdot X^T + \lambda M \sigma^2)(X \cdot X^T + \lambda\sigma^2 I)^{-1}$

If M is a zero matrix, $A_{MAP} = (Y \cdot X^T)(X \cdot X^T + \lambda\sigma^2 I)^{-1}$

Name: Xun Xue
UNI: xx2241

(V) 1. The result in (i) and (iii) are the same

$$P((X,Y)|A) = P((x^1,y^1),(x^2,y^2),\dots(x^m,y^m)|A) = \prod_{i=1}^{m} P(x_i|A)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{|\sigma^2 I|^{\frac{1}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)\right)$$

After we calculate the logrithm of $P((X,Y)|A)$, the crucial part of the result is $\sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)$, which is exactly L_{LS}.

Since when L_{LS} is the minimum, $P((X,Y)|A)$ is the maximum, so (i) and (iii) is the same problem, therfore the result is the same.

2. The result in (ii) and (iv) are similar

In (iv) $A_{MAP} = (Y \cdot X^T + \lambda M \sigma^2)(X \cdot X^T + \lambda \sigma^2 I)^{-1}$

if M is a zero matrix and $\sigma^2 = 1$, then the result of A_{MAP} in (iv) is equal to A_r in (ii).

When we calculate A_{MAP}, we calculate $\frac{d \log P((X,Y)|A)}{dA}$ and $\frac{d \log P(A)}{dA}$

notice that $\frac{d \log(P((X,Y)|A))}{dA}$ is same as $\frac{d\left(\sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)\right)}{dA}$

the crucial part of $\log P(A)$ is $Tr[\lambda(A-M)^T(A-M)]$, when M is a zero matrix, this term is $\lambda Tr(A^TA)$. Since $\|A\|_F^2 = Tr(A^TA)$, this term is exactly as $\lambda\|A\|_F^2$ in L_r. Therefore when $\sigma^2 = 1$ and M is a zero matrix result in (ii) and (iv) will be the same.