ECBME 4040 Homework 2 Writeup

Jingyu Qian (UNI jq2250)

i

$A_{ls} = \arg\min_A L_{ls}$ infers we found the derivative $w.r.t$ A such that the derivative reaches 0.

$$\begin{aligned}
&\nabla_A \sum_{i=1}^{m} (y^i - Ax^i)^T (y^i - Ax^i) \\
&= \nabla_A \sum_{i=1}^{m} (y^{iT} y^i - x^{iT} A^T y^i - y^{iT} A x^i + x^{iT} A^T A x^i) \\
&= \sum_{i=1}^{m} -2y^i x^{iT} + 2Axx^{iT} \\
&= -2YX^T + 2AXX^T = 0
\end{aligned}$$

$$AXX^T = YX^T$$
$$A = YX^T(XX^T)^{-1} = YX^{-1}$$

ii

Similar to question i, this time:

$$AXX^T + \lambda AI = YX^T$$
$$A = YX^T(XX^T + \lambda I)^{-1}$$

iii

$$\begin{aligned}
A_{ML} &= \arg\max_A \prod_{i=1}^{m} P(\epsilon_i | A) \\
&= \arg\max_A \sum_{i=1}^{m} \ln(P(\epsilon_i | A))
\end{aligned}$$

$$\begin{aligned}
P(\epsilon_i; A) &= N(0, \sigma^2 I) \\
&= \frac{1}{\sqrt{(2\pi)^n \sigma^2 I}} e^{-\frac{1}{2}(y^i - Ax^i)^T \sigma^{-2} I (y^i - Ax^i)}
\end{aligned}$$

$$
\begin{aligned}
\therefore A_{ML} &= \arg\max_A -\frac{1}{2}\sum_{i=1}^{m}(y^i - Ax^i)^T\sigma^{-2}I(y^i - Ax^i) \\
&= \arg\min_A \sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)
\end{aligned}
$$

It is obvious the answer should be the same as problem i. So the answer is:

$$A = YX^{-1}$$

iv

$$A_{MAP} = \arg\max_A P(\epsilon|A)P(A)$$

$$P(A) = MN(M, \lambda^{-\frac{1}{2}}I, \lambda^{-\frac{1}{2}}I)$$

$$
\begin{aligned}
\therefore A_{MAP} &= \arg\max_A \prod P(\epsilon_i|A) \cdot P(A) \\
&= \arg\max_A \sum \ln P(\epsilon_i|A) + \ln P(A) \\
&= \arg\max_A (-\frac{1}{2}\sum_{i=1}^{m}(y^i - Ax^i)^T\sigma^{-2}I(y^i - Ax^i) - \frac{\lambda}{2}tr[I(A-M)^T I(A-M)]) \\
&= \arg\max_A (-\sigma^{-2}\sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i) - \lambda \cdot tr[(A-M)^T(A-M)])
\end{aligned}
$$

Still we can calculate the gradient w.r.t A:

$$
\begin{aligned}
&\nabla_A(-\sigma^{-2}\sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i) - \lambda \cdot tr[(A-M)^T(A-M)]) \\
&= -\sigma^{-2}(-2YX^T + 2AXX^T) - 2\lambda \cdot (A - M) = 0
\end{aligned}
$$

$$A = (\sigma^{-2}YX^T + \lambda M)(\sigma^{-2}XX^T + \lambda I)^{-1}$$

v

Result **i** and **iii** are the same, this means we can use maximum likelihood estimate as an unbiased estimate of the least square loss function. Result **ii** and **iv** are similar in their forms. So given A's priori we can also use MAP estimate in place of the least square loss function with Frobenius regularization form.