ECBM: 6040 : Neural Networks and Deep Learning

Homework #2

Name: Siddharth Aman Varshney Uni: sav2125

Problem 2: $Y = [y^1\ y^2 \ldots\ y^m]$ $X = [x^1\ x^2 \ldots\ x^m]$

(i) $\mathcal{L}_{LS} = \sum_{i=1}^{m} (y^i - Ax^i)^T (y^i - Ax^i)$

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

$= \sum_{i=1}^{m} \left[(y^i)^T y^i - (y^i)^T A x^i - (x^i)^T A^T y^i + (x^i)^T A^T A x^i \right]$

$A_{LS} = \underset{A}{\arg\min}\ \mathcal{L}_{LS}$

To find A_{LS}, we differentiate $\mathcal{L}_{LS}$ w.r.t. A

$\frac{\partial \mathcal{L}_{LS}}{\partial A} = \frac{\partial}{\partial A}\left[\sum_{i=1}^{m} \left((y^i)^T y^i - (y^i)^T A x^i - (x^i)^T A^T y^i + (x^i)^T A^T A x^i \right) \right]$

Now $\frac{\partial(a^T X b)}{\partial X} = ab^T$, $\frac{\partial(a^T X^T b)}{\partial X} = ba^T$

$\therefore$ and $\frac{\partial(b^T X^T X c)}{\partial X} = X\{bc^T + cb^T\}$

Thus $\frac{\partial(\mathcal{L}_{LS})}{\partial A} = \sum_{i=1}^{m} -2y^i(x^i)^T + 2A(x^i(x^i)^T) = 0$

$\Rightarrow \sum_{i=1}^{m} y^i(x^i)^T = A\sum_{i=1}^{m}(x^i(x^i)^T)$

$\Rightarrow A = \left(\sum_{i=1}^{m} y^i(x^i)^T\right)\left(\sum_{i=1}^{m} x^i(x^i)^T\right)^{-1}$

(ii) $\mathcal{L}_r = \lambda\|A\|_F^2 + \sum_{i=1}^{m}(y^i - Ax^i)^T(y^i - Ax^i)$

$A_r = \underset{A}{\arg\min}\ \mathcal{L}_r$

Now $\|A\|_F^2 = Tr(A^T A)$

To find A_r, take gradient of $\mathcal{L}_r$ w.r.t. A

$$\therefore \frac{\partial \alpha_r}{\partial A} = \frac{\partial}{\partial A}\left[\lambda \|A\|_F^2 + \sum_{i=1}^{m} (y^i - Ax^i)^T (y^i - Ax^i) \right]$$

$$\frac{\partial \alpha_r}{\partial A} = 2\lambda A - 2\sum_{i=1}^{m} y_i^i (x^i)^T + 2A \sum_{i=1}^{m} (x^i)(x^i)^T = 0$$

$$\Rightarrow \quad A\left[\lambda I + \sum_{i=1}^{m} (x^i)(x^i)^T\right] = \sum_{i=1}^{m} y^i (x^i)^T$$

$$\Rightarrow \quad A_n = \left(\sum_{i=1}^{m} y^i (x^i)^T\right)\left(\lambda I + \sum_{i=1}^{m} (x^i)(x^i)^T\right)^{-1}$$

(iii) $\varepsilon_i = y^i - Ax^i$, $\varepsilon_i \sim N(0, \sigma^2 I)$

Since we need to find A_{ML},

Maximum likelihood formula is

$$l = \sum_{i=1}^{m} \log \frac{1}{(2\pi)^{n/2}\sqrt{|\sigma^2 I|}} \times \exp\left(-\frac{1}{2}(e_i)^T (\sigma^2 I)^{-1} e_i\right)$$

$$\Rightarrow l = -\frac{mn}{2}\log 2\pi + \frac{m}{2}\log|(\sigma^2 I)^{-1}| - \frac{1}{2}\left[(e_i)^T (\sigma^2 I)^{-1} e_i\right]$$

Scanned by CamScanner

$$\frac{\partial l}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{1}{2}\Sigma (y^i - Ax^i)^T (\sigma^2 I)^{-1} (y^i - Ax^i) \right]$$

Now $\quad x^T A x = tr(x^T A x]$

$$\therefore \frac{\partial l}{\partial A} = \frac{\partial}{\partial A}\left[-\frac{1}{2} \cdot \sum_{i=1}^{m} tr\left((y^i - Ax^i)^T (\sigma^2 I)^{-1} (y^i - Ax^i) \right) \right]$$

$$= \frac{\partial}{\partial A}\left[\frac{-\sigma^2}{2} \cdot \sum_{i=1}^{m} tr (y^i - Ax^i)^T (y^i - Ax^i) \right]$$

$$= \frac{\partial}{\partial A}\left[-\frac{(\sigma^2)^{-1}}{2} \times \sum_{i=1}^{m} tr\left[(y^i)^T y^i - (y^i)^T A x^i - (x^i)^T A^T y^i + (x^i)^T A^T A x^i \right] \right]$$

$$= \frac{(\sigma^2)^{-1}}{2} \sum_{i=1}^{m} -(y^i (x^i)^T) + A x^i (x^i)^T) = 0$$

$$\left[\because \; tr(x^T A x) = x^T A x \right]$$

$$\Rightarrow \quad A = \left(\sum_{i=1}^{m} (y^i)(x^i)^T \right) \left(\sum_{i=1}^{m} (x^i)(x^i)^T \right)^{-1}$$

(iv)

$A \sim MN(M, \lambda^{-1/2}, I, \lambda^{-1/2} I)$

To find A_{MAP},

$$A_{MAP} = \arg\max_A \log \prod_{i=1}^{m} p(\varepsilon_i|A)\, p(A)$$

$$l = \log \prod_{i=1}^{m} p(\varepsilon_i|A)\, p(A)$$

$$\Rightarrow l = \sum_{i=1}^{m} \log \left\{ \frac{1}{(2\pi)^{n/2} \sqrt{|\sigma^2 I|}} \cdot \exp\left(\frac{-1}{2} (e_i)^T (\sigma^2 I)^{-1} e_i\right) \times \frac{1}{(2\pi)^{n^2/2}} \exp\left(\frac{-1}{2} \cdot Tr[\lambda (A-M)^T (A-M)]\right) \right\}$$

$$\Rightarrow l = \sum_{i=1}^{m} -\frac{mn}{2} \log 2\pi + \frac{m}{2} \log |(\sigma^2 I)^{-1}| - \frac{1}{2} [(y^i - Ax^i)^T (\sigma^2 I)^{-1} (y^i - Ax^i)$$

$$+ \log\left(\frac{1}{(2\pi)^{n^2/2}}\right) - \frac{1}{2} \times Tr(\lambda (A-M)^T (A-M))$$

Scanned by CamScanner

To find A_{MAP}, take derivative of l w.r.t. A.

$$\frac{\partial l}{\partial A} = (\sigma^2)^{-1} \sum_{i=1}^{m} \left\{ (y^i)(x^i)^T - A x^i (x^i)^T \right\} + \lambda (M - A) = 0$$

$$\left[\because \frac{\partial\, Tr(AX^T)}{\partial X} = A \;,\; \frac{\partial}{\partial X} Tr(X^T A) = A \right]$$

$$\Rightarrow \quad \cancel{(\sigma^2)^{-1} \sum \lambda M}$$

$$(\sigma^2)^{-1} \sum_{i=1}^{m} (y^i)(x^i)^T + \lambda M = A\left[\lambda I + (\sigma^2)^{-1} \sum_{i=1}^{m} x^i (x^i)^T \right]$$

$$\Rightarrow \quad A_{MAP} = \left(\left(\sum_{i=1}^{m} y^i (x^i)^T \right) + \sigma^2 \lambda M \right) \left(\sigma^2 \lambda I + \sum_{i=1}^{m} x^i (x^i)^T \right)^{-1}$$

If $M = 0$, then

$$A_{MAP} = \left(\sum_{i=1}^{m} y^i (x^i)^T \right) \left(\sigma^2 \lambda I + \sum_{i=1}^{m} x^i (x^i)^T \right)^{-1}$$

(v) $A_{LS} = \underset{A}{\arg\min}\, \mathcal{L}_{LS}$ is essentially the same expression as A_{ML}, where A_{ML} is the maximum likelihood estimate of A.

A_{LS} is derived in (i), A_{ML} in (iii)

Now after calculation we get

$$A_R = \left(\sum_{i=1}^{m} y^i (x^i)^T \right) \left(\lambda I + \sum_{i=1}^{m} (x^i)(x^i)^T \right)^{-1}$$

$$A_{MAP} = \left(\sum_{i=1}^{M} y^i (x^i)^T + \sigma^2 \lambda M \right) \left(\sigma^2 \lambda I + \sum_{i=1}^{m} x^i (x^i)^T \right)^{-1}$$

$A_R = A_{MAP}$ when $M = 0$, $\sigma = 1$

or when $\sigma = 0$