Zoltan Onodi-Szucs

zo213

Deep Learning

Homework II

Part I - Theory

Problem i,

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

$$= \mathrm{Tr}\left[(Y - AX)^T (Y - AX)\right]$$

$$\nabla_A \mathcal{L}_{ls} = \nabla_A \mathrm{Tr}\left[\left(Y^T Y - 2X^T A^T Y + X^T A^T A X\right)\right]$$

$$= -2YX^T + 2A(XX^T)$$

$$\Downarrow$$

$$0 = -2YX^T + 2A(XX^T)$$

$$A_{ls} = (YX^T) \cdot (XX^T)^{-1}$$

Problem ii,

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

$$= \lambda \, Tr(A^T A) + Tr\left[(Y - AX)^T (Y - AX)\right]$$

$$\nabla_A \mathcal{L}_r = \lambda 2A - 2YX^T + 2AXX^T$$

$$\Downarrow$$

$$0 = \lambda A - YX^T + AXX^T$$

$$A_r = YX^T \cdot (\lambda I + XX^T)^{-1}$$

Problem iii,

$$\varepsilon_i = y^i - Ax^i \;, \; \varepsilon_i \sim \mathcal{N}(0, \sigma^2 I)$$

$$P(\varepsilon \mid 0, \sigma^2 I) = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2} \frac{\varepsilon_i^2}{\sigma^2}\right\}$$

$$\mathcal{L} = -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\varepsilon_i$$

$$= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\mathrm{Tr}\left[(Y - AX)^T(Y - AX)\right]$$

$$\nabla_A \mathcal{L} = -\frac{1}{2\sigma^2}\cdot\left(-2YX^T + AXX^T\right)$$

$$A_{ML} = YX^T\cdot(XX^T)^{-1}$$

Problem iv,

$$P(A|\varepsilon) = P(\varepsilon|A)\cdot P(A)$$

$$= \prod_{i=1}^{m}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{\varepsilon_i^2}{\sigma^2}\right\}\cdot\frac{1}{(2\pi)^{n^2/2}}\exp\left\{-\frac{1}{2}\mathrm{Tr}(\lambda(A-M)^T(A-M))\right.$$

$$\mathcal{L} = -\frac{m}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\varepsilon_i^2 - \frac{m^2}{2}\ln 2\pi - \frac{1}{2}\mathrm{Tr}(\lambda(A-M)^T(A-M))$$

$$= -\frac{1}{2}\mathrm{Tr}\left[(Y - AX)^T(Y - AX)\right] - \frac{1}{2}\mathrm{Tr}(\lambda(A-M)^T(A-M)) + \text{const w.r.t. } A$$

$$\nabla_A \mathcal{L} = -\frac{1}{2\sigma^2}\left(-2YX^T + 2AXX^T\right) - \frac{1}{2}\left(2\lambda A - 2M\right)$$

$$\Downarrow$$

$$0 = -\frac{1}{2\sigma^2}\left(-2YX^T + 2AXX^T\right) - \frac{1}{2}\left(2\lambda A - 2M\right)$$

$$YX^T = AXX^T + \sigma^2(\lambda A - M)$$

$$\Big\downarrow \; M = 0$$

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

Problem v,

The estimator for A in problem i and iii is the same, because we made an independent Gaussian noise assumption about the error. The least square solution is equal to the maximum likelihood one.

The estimators in the case of problem ii and iv also has the same structure. The MAP and R solution is the same (modulo the shift from λ to $\sigma^2\lambda$). The prior on A acts as the regulating term in the case of the R estimate. Maximizes the posterior.