Zoltan Onodi-Szucs
zo213

Deep Learning
Homework I

Part II - Theory

Problem c

Problem C

$$P_x(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{else} \end{cases}$$

$$y = -\frac{1}{\lambda} \ln(x)$$

$$-\lambda y = \ln(x)$$

$$e^{-\lambda y} = x \qquad \frac{dx}{dy} = -\lambda e^{-\lambda y}$$

$$-\lambda y \leq 0$$
$$y \geq 0$$
$$\lambda > 0$$
$$\Downarrow$$

$$P_y(y) = P_x(x(y)) \cdot \left|\frac{dx}{dy}\right| = 1 \cdot \left|-\lambda e^{-\lambda y}\right| = \left|-\lambda e^{-\lambda y}\right| = \lambda e^{-\lambda y}$$

$$\int_{-\infty}^{\infty} P(y)\, dy = \int_0^{\infty} \lambda e^{-\lambda y}\, dy =$$

$$= \left[-e^{-\lambda y}\right]_0^{\infty} = \left[-\frac{1}{e^{\lambda y}}\right]_0^{\infty} =$$

$$= 0 + 1 = 1$$

ii)

$$P(X=x, Y=y) = \begin{cases} 3(xy^2+yx^2) & , x,y \in [0,1] \\ 0 & , \text{else} \end{cases}$$

$$P(X=x) = \int_0^1 p(x,y)\,dy = 3\left[\frac{xy^3}{3} + \frac{y^2x^2}{2}\right]_0^1 = x + \frac{3}{2}x^2$$

$$P(Y=y) = \int_0^1 p(x,y)\,dx = 3\left[\frac{x^2y^2}{2} + \frac{yx^3}{3}\right]_0^1 = \frac{3}{2}y^2 + y$$

$$E(x) = \int_0^1 p(x)\cdot x\,dx = \int_0^1 (x^2 + \frac{3}{2}x^3)\,dx = \left[\frac{x^3}{3} + \frac{3}{8}x^4\right]_0^1 = \frac{1}{3} + \frac{3}{8} = \frac{17}{24}$$

$$E(y) = \int_0^1 p(y)\cdot y\,dy = \int_0^1 (y^2 + \frac{3}{2}y^3)\,dy = \left[\frac{y^3}{3} + \frac{3}{8}y^4\right]_0^1 = \frac{17}{24}$$

$$E(xy) = \int_x\int_y 3(xy^2+yx^2)\cdot x\cdot y\,dy\,dx = 3\int_x\int_y (x^2y^3+y^2x^3)\,dy\,dx =$$

$$= 3\int_x \left[\frac{x^2y^4}{4} + \frac{y^3x^3}{3}\right]_0^1 dx = 3\int_x \left(\frac{x^2}{4} + \frac{x^3}{3}\right)dx = 3\left[\frac{x^3}{12} + \frac{x^4}{12}\right]_0^1 =$$

$$= \frac{1}{2}$$

$E(xy) \neq E(x)\cdot E(y) \Rightarrow$ dependent

Problem d

Problem d

i) $X \sim N(\mu, \Sigma) \quad \theta = \{\mu, \Sigma\}$

$$\hat{\theta}_{ML} = \arg\max_{\theta} \prod_{i=1}^{n} p(x_i|\theta) = \arg\max_{\theta} \sum_{i=1}^{n} \ln p(x_i|\theta)$$

$$0 = \nabla_{\mu} \sum_{i=1}^{n} \ln \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x_i-\mu)^T \Sigma^{-1} (x_i-\mu)\right)$$

$$= \nabla_{\mu} \sum_{i=1}^{n} -\frac{1}{2} \ln (2\pi)^n |\Sigma| - \frac{1}{2} (x_i-\mu) \Sigma^{-1} (x_i-\mu)$$

$$= -\frac{1}{2} \sum_{i=1}^{n} \nabla_{\mu} \left(x_i^T \Sigma^{-1} x_i - 2\mu^T \Sigma x_i + \mu^T \Sigma^{-1} \mu\right) =$$

$$= -\Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu)$$

Σ is positive definite

$\Downarrow$

Only solution: $\sum_{i=1}^{n} (x_i - \mu) = 0$

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$0 = \nabla_\Sigma \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi)^d|\Sigma| - \frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)$$

$$= -\frac{n}{2}\nabla_\Sigma \ln|\Sigma| - \frac{1}{2}\nabla_\Sigma \operatorname{trace}\left(\Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T\right)$$

$$= -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-2}\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$$

$$\Downarrow$$

$$\hat{\Sigma}_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu}_{ML})(x_i-\hat{\mu}_{ML})^T$$

ii)

$$\text{bias}(\hat{\mu}) = E(\hat{\mu}) - \mu = E\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right) - \mu =$$

$$= \frac{1}{n}E(\sum_{i=1}^{n}x_i - \mu = 0$$

$\Rightarrow$ unbiased

$$E(\hat{\Sigma}) = E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu}_{ML})(x_i - \hat{\mu}_{ML})^T\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i)^2 - E(\hat{\mu}_{ML})^2 = \frac{1}{n}\sum_{i=1}^{n} E(x_i)^2 - E\,\frac{1}{n}\sum_{i=1}^{n} x_i\,\frac{1}{n}\sum_{j=1}^{n} x_j =$$

$$= \frac{1}{n}\left(1 - \frac{1}{n}\right)\sum_{i=1}^{n} E(x_i)^2 - \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i} E\,x_i x_j =$$

$$= \frac{1}{n}\left(1 - \frac{1}{n}\right)n\left(\hat{\Sigma} + \mu^2\right) - \frac{1}{n^2}n(n-1)\mu^2 = \frac{n-1}{n}\Sigma$$

$$\text{bias}(\hat{\Sigma}) = \frac{n-1}{n}\Sigma - \Sigma = \frac{1}{n}\Sigma \Rightarrow \text{biased}$$