DeepLearning HW1

Sirui Tan

February 14, 2016

1 Problem a

See `hw1a.py`

2 Problem b

See `hw1b.py`

3 Problem c

3.1 i

Let $y = g(x)$, since g is monotonic, we have

$$f_Y(y) = \left| \frac{d}{dy}(g^{-1}(y)) \right| \cdot f_X(g^{-1}(y))$$

Take $g(x) = -\frac{1}{\lambda}\ln(x)$ into the formula

$$f_Y(y) = |\lambda| e^{-\lambda y} f_X(e^{-\lambda y})$$

If $\lambda > 0$:

$$f_Y(y) = \begin{cases} \lambda e^{-y} & y \geq 0 \\ 0 & otherwise \end{cases}$$

else

$$f_Y(y) = \begin{cases} -\lambda e^{-y} & y \leq 0 \\ 0 & otherwise \end{cases}$$

3.2 ii

$$\begin{aligned} p(X=x) &= \int p(X=x, Y=y)dy \\ &= \int_0^1 3(xy^2 + yx^2)dy \\ &= x + frac3x^2 2, 0 \leq x \leq 1 \\ p(Y=y) &= \int p(X=x, Y=y)dx \\ &= \int_0^1 3(xy^2 + yx^2)dx \\ &= y + frac3y^2 2, 0 \leq y \leq 1 \\ E(x) &= \int x \cdot p(x)dx \\ &= \int_0^1 x^2 + \frac{3}{2}x^3 dx \\ &= \frac{17}{24} \\ E(y) &= \int y \cdot p(y)dy \\ &= \int_0^1 y^2 + \frac{3}{2}y^3 dy \\ &= \frac{17}{24} \\ E(xy) &= \int_0^1 \int_0^1 xy \cdot 3(xy^2 + yx^2)dxdy \\ &= \frac{1}{2} \end{aligned}$$

4 Problem d

4.1 i

Likelihood:

$$\begin{aligned} p(X|\mu, \Sigma) &= p(x_1, x_2, \cdots, x_m|\mu, \Sigma) \\ &= \prod_{i=1}^{m} p(x_i|\mu, \Sigma) \end{aligned}$$

Convert to log-likelihood:

$$L(X|\mu, \Sigma) = \sum_{i=1}^{m} \log(\frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}}) \exp(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1}(x_i - \mu))$$

Optimize L over μ is equivalent to solve the equation:

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{m}-\frac{1}{2}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu)=0$$

which can be simplified as:

$$\frac{\partial}{\partial\mu}\sum_{i=1}^{m}-\frac{1}{2}(x_i^T\Sigma^{-1}x_i-2x_i^T\Sigma^{-1}\mu+\mu^T\Sigma^{-1}\mu)=0$$

Using relationship

$$\frac{\partial X^TAX}{\partial X}=X^T(A+A^T)$$

The equation can be further simplified as

$$\sum_{i=1}^{m}(x_i^T-\mu^T)\Sigma^{-1}=0$$

So

$$\mu=\frac{\sum_{i=1}^{m}x_i}{m}$$

Now calculate Σ. Simplify again over log-likelihood w.r.t Σ:

$$L=-\frac{mn}{2}\log 2\pi+\frac{m}{2}\log|A|-\frac{1}{2}\sum_{i=1}^{m}tr[(x-\mu)(x-\mu)^TA]$$

Using relationship

$$\frac{\partial\log|A|}{\partial A}=(A^{-1})^T$$
$$\frac{\partial\log|BA|}{\partial A}=B^T$$

to derive the derivative w.r.t Σ

$$\frac{\partial L}{\partial\Sigma}=\frac{m}{2}\Sigma-\frac{1}{2}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T=0$$

Thus

$$\Sigma=\frac{1}{m}\sum_{i=1}^{m}(x-\mu)(x-\mu)^T$$

4.2 ii

Check the biasness of μ_{ML}:

$$\begin{aligned}
E(\mu_{ML}) - \mu &= \frac{1}{m}\sum_{n=1}^{m} E(x_i) - \mu \\
&= \frac{1}{m}\sum_{n=1}^{m} \mu - \mu \\
&= 0
\end{aligned}$$

Therefore, μ_{ML} is unbiased.

Now check the biasness of Σ_{ML}:

$$\begin{aligned}
E(\Sigma_{ML}) &= \frac{1}{m}\sum_{i=1}^{m} E[(x_i - \frac{1}{m}\sum_{j=1}^{m} x_j)(x_i^T - \frac{1}{m}\sum_{j=1}^{m} x_j^T)] \\
&= \frac{1}{m}\left\{\sum_{i=1}^{m} E[x_i x_i^T] - \sum_{i=1}^{m} E[\frac{1}{m} x_i \sum_{k=1}^{m} x_k^T] - \sum_{i=1}^{m} E[\frac{1}{m}\sum_{j=1}^{m} x_j x_i^T] + \sum_{i=1}^{m} E[\frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m} x_j x_k^T]\right\} \\
&= \frac{1}{m}\left\{m(\mu\mu^T + \Sigma) - (m\mu\mu^T + \Sigma)\right\} \\
&= \frac{m-1}{m}\Sigma
\end{aligned}$$

Clearly, Σ_{ML} is biased.