CS405 · Machine Learning

Assignment 1

何泽安 (He Zean)
12011323
Sep. 24, 2022

Question 1

$$
\begin{aligned}
E(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\left[y(x_n, \mathbf{w}) - t_n\right]^2 \\
&= \frac{1}{2}\sum_{n=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n - t_n\right)^2
\end{aligned}
$$

w.r.t $\mathbf{x}_n = \begin{bmatrix} x_n^0 & \cdots & x_n^M \end{bmatrix}^{\mathrm{T}}$, n is the index of sample. To minimize the error function, find its partial derivative about $\mathbf{w}$.

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w}}E(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}\left[\frac{\partial}{\partial \mathbf{w}}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n - t_n\right)^2\right] \\
&= \frac{1}{2}\sum_{n=1}^{N}\left\{2\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n - t_n\right)\begin{bmatrix} \frac{\partial}{\partial \mathbf{w}_0}\left(w_0 x_n^0 + \cdots + w_M x_n^M - t_n\right) \\ \vdots \\ \frac{\partial}{\partial \mathbf{w}_M}\left(w_0 x_n^0 + \cdots + w_M x_n^M - t_n\right) \end{bmatrix}\right\} \\
&= \frac{1}{2}\sum_{n=1}^{N}\left\{2\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n - t_n\right)\begin{bmatrix} x_n^0 \\ \vdots \\ x_n^M \end{bmatrix}\right\} \\
&= \sum_{n=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_n - t_n\right)\mathbf{x}_n = \mathbf{0}
\end{aligned}
$$

We can figure out that for the i^{th} row in the column vector, such equation establishes:

$$\textstyle\sum_{m=0}^{M}\sum_{n=0}^{N} w_m x_n^m x_n^i = \sum_{n=0}^{N} t_n x_n^i$$

Denoting $X = \begin{bmatrix} \sum_n x_n^{(i+j)} \end{bmatrix}_{M\times M}$ where i, j is the row/col index of matrix element; $Y = \begin{bmatrix} \sum_n t_n x_n^i \end{bmatrix}_{M\times 1}$:

$$X\mathbf{w} = Y$$

$$X^T X\mathbf{w} = X^T Y \Rightarrow \mathbf{w} = (X^T X)^{-1} X^T Y$$

Question 2

$$
\begin{aligned}
p(\text{apple}) &= p(\text{apple}|r)\cdot p(r) + p(\text{apple}|g)\cdot p(g) + p(\text{apple}|b)\cdot p(b) \\
&= \frac{3}{3+4+3}\cdot 0.2 + \frac{3}{3+3+4}\cdot 0.6 + \frac{1}{1+1+0}\cdot 0.2 \\
&= 0.34
\end{aligned}
$$

$$
\begin{aligned}
p(g|\text{orange}) &= \frac{p(\text{orange}|g)\cdot p(g)}{p(\text{orange})} \\
&= \frac{p(\text{orange}|g)\cdot p(g)}{p(\text{orange}|r)\cdot p(r) + p(\text{orange}|g)\cdot p(g) + p(\text{orange}|b)\cdot p(b))} \\
&= \frac{3/(3+3+4)\cdot 0.6}{4/(3+4+3)\cdot 0.2 + 3/(3+3+4)\cdot 0.6 + 1/(1+1+0)\cdot 0.2} \\
&= 0.5
\end{aligned}
$$

Question 3

$$
\begin{aligned}
\mathbb{E}[x+z] &= \iint (x+z)p(x,z)\,\mathrm{d}x\,\mathrm{d}z \\
&\overset{\text{indep.}}{\underset{x,z}{=\!=\!=}} \iint (x+z)p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z \\
&= \iint xp(x)p(z)\,\mathrm{d}x\,\mathrm{d}z + \iint zp(z)p(x)\,\mathrm{d}x\,\mathrm{d}z \\
&= \int \left[\int p(z)\,\mathrm{d}z\right] xp(x)\,\mathrm{d}x + \int \left[\int p(x)\,\mathrm{d}x\right] zp(z)\,\mathrm{d}z \\
&= \int 1\cdot xp(x)\,\mathrm{d}x + \int 1\cdot zp(z)\,\mathrm{d}z \\
&= \mathbb{E}[x] + \mathbb{E}[z]
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{var}[x] + \mathbf{var}[z] &= \int \{x-\mathbb{E}[x]\}^2 p(x)\,\mathrm{d}x + \int \{z-\mathbb{E}[z]\}^2 p(z)\,\mathrm{d}z \\
&\overset{\int p(x)\,\mathrm{d}x=1}{\underset{\int p(z)\,\mathrm{d}z=1}{=\!=\!=\!=}} \int \left\{\int p(z)\,\mathrm{d}z\right\} \{x-\mathbb{E}[x]\}^2 p(x)\,\mathrm{d}x + \int \left\{\int p(x)\,\mathrm{d}x\right\} \{z-\mathbb{E}[z]\}^2 p(z)\,\mathrm{d}z \\
&= \iint \{x-\mathbb{E}[x]\}^2 p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z + \iint \{z-\mathbb{E}[z]\}^2 p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z \\
&= \iint \left\{\{(x+z)-\mathbb{E}[x+z]\}^2 - 2xz - 2\mathbb{E}[x]\mathbb{E}[z] + 2x\mathbb{E}[z] + 2z\mathbb{E}[x]\right\} p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z \\
&= \mathbf{var}[x+z] - 2\iint \{xz + \mathbb{E}[x]\mathbb{E}[z] - x\mathbb{E}[z] - z\mathbb{E}[x]p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z\} \\
&= \mathbf{var}[x+z] - 2\{\int \left[\int zp(z)\,\mathrm{d}z\right] xp(x)\,\mathrm{d}x + \mathbb{E}[x]\mathbb{E}[z]\iint p(x)p(z)\,\mathrm{d}x\,\mathrm{d}z \\
&\qquad - \mathbb{E}[z]\int \left[\int xp(x)\,\mathrm{d}x\right] p(z)\,\mathrm{d}z - \mathbb{E}[x]\int \left[\int zp(z)\,\mathrm{d}z\right] p(x)\,\mathrm{d}x\} \\
&= \mathbf{var}[x+z] - 2\{\mathbb{E}[x]\mathbb{E}[z] + \mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[x]\mathbb{E}[z]\} \\
&= \mathbf{var}[x+z]
\end{aligned}
$$

Question 4

Poisson Distribution Having its likelihood function, consider the partial derivation about λ of its log-likelihood:

$$
\begin{gathered}
p(\mathcal{D}|\lambda) = \textstyle\prod_{i=1}^{n} \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} \\
\ln p(\mathcal{D}|\lambda) = \textstyle\sum_{i=1}^{n} \ln \frac{\lambda^{X_i} e^{-\lambda}}{X_i!} = \sum_{i=1}^{n} [X_i \ln\lambda - \lambda - \ln(X_i!)] \\
\frac{\partial}{\partial\lambda} \ln p(\mathcal{D}|\lambda) = \frac{\sum_{i=1}^{n} X_i}{\lambda} - n = 0 \Rightarrow \lambda_{\mathrm{ML}} = \frac{1}{n}\textstyle\sum_{i=1}^{n} X_i
\end{gathered}
$$

Exponential Distribution By the definition of exponential distribution, all the sample points in $\mathcal{D}$ are greater than 0 (otherwise, $p(\mathcal{D}|\lambda)$ gotta be 0):

$$p(\mathcal{D}|\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{-X_i/\lambda}$$
$$\ln p(\mathcal{D}|\lambda) = \sum_{i=1}^{n} \ln\left(\frac{1}{\lambda} e^{-X_i/\lambda}\right) = \sum_{i=1}^{n}\left(-\frac{X_i}{\lambda} - \ln\lambda\right)$$
$$\frac{\partial}{\partial\lambda}\ln p(\mathcal{D}|\lambda) = \frac{\sum_{i=1}^{n} X_i}{\lambda^2} - \frac{n}{\lambda} = 0 \Rightarrow \lambda_{\text{ML}} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Question 5

$p(\text{correct}) = p(x \in \mathcal{R}_1, C_1) + p(x \in \mathcal{R}_2, C_2) = \int_{\mathcal{R}_1} p(x, C_1)\,\mathrm{d}x + \int_{\mathcal{R}_2} p(x, C_2)\,\mathrm{d}x$
$p(\text{mistake}) = p(x \in \mathcal{R}_2, C_1) + p(x \in \mathcal{R}_1, C_2) = \int_{\mathcal{R}_2} p(x, C_1)\,\mathrm{d}x + \int_{\mathcal{R}_1} p(x, C_2)\,\mathrm{d}x$

The calculus of variations' way can be repretated as:

$$\frac{\delta\mathbb{E}[L]}{\delta y(\mathbf{x})} = 2\int \{y(\mathbf{x}) - \mathbf{t}\}\, p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t} = 0$$
$$y(\mathbf{x})\int p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t} = \int \mathbf{t}p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t}$$
$$y(\mathbf{x}) = \frac{\int \mathbf{t}p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t}}{\int p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t}} = \frac{\int \mathbf{t}p(\mathbf{x}, \mathbf{t})\,\mathrm{d}\mathbf{t}}{p(\mathbf{x})} = \int \mathbf{t}p(\mathbf{t}|\mathbf{x})\,\mathrm{d}\mathbf{t} = \mathbb{E}_{\mathbf{t}}[\mathbf{t}|\mathbf{x}]$$

Question 6

$$\begin{aligned}
\mathbf{H}[\mathbf{X}] &= -\int p(x)\ln\left[\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right]\mathrm{d}x \\
&= \int \ln(\sqrt{2\pi}\sigma)p(x) + \frac{(x-\mu)^2}{2\sigma^2}p(x)\,\mathrm{d}x \\
&= \ln(\sqrt{2\pi}\sigma)\int p(x)\,\mathrm{d}x + \frac{1}{2\sigma^2}\left[\int x^2 p(x)\,\mathrm{d}x - 2\mu\int xp(x)\,\mathrm{d}x + \mu^2\int p(x)\,\mathrm{d}x\right] \\
&= \ln(\sqrt{2\pi}\sigma)\cdot 1 + \frac{1}{2\sigma^2}\left\{\mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mu^2\cdot 1\right\} \\
&= \ln(\sqrt{2\pi}\sigma)\cdot 1 + \frac{1}{2\sigma^2}\left\{\mathbf{var}[x] + \mathbb{E}^2[x] - 2\mu\mathbb{E}[x] + \mu^2\right\} \\
&\xlongequal{x\sim\mathcal{N}(\mu,\sigma^2)} \ln(\sqrt{2\pi}\sigma) + \frac{1}{2\sigma^2}\left\{\sigma^2 + \mu^2 - 2\mu\cdot\mu + \mu^2\right\} \\
&= \ln(\sqrt{2\pi}\sigma) + \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
\mathbf{I}[\mathbf{x}, \mathbf{y}] &\equiv \mathbf{KL}(p(\mathbf{x}, \mathbf{y})\|p(\mathbf{x})p(\mathbf{y})) \\
&= -\iint p(\mathbf{x}, \mathbf{y})\ln\left(\frac{p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x}, \mathbf{y})}\right)\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

$$
\begin{aligned}
\mathbf{I}[\mathbf{x},\mathbf{y}] &= -\iint p(\mathbf{x},\mathbf{y}) \ln\left(\frac{p(\mathbf{x})}{p(\mathbf{x},\mathbf{y})/p(\mathbf{y})}\right) \mathrm{d}x\,\mathrm{d}y \\
&= -\iint p(\mathbf{x},\mathbf{y}) \left[\ln p(\mathbf{x}) - \ln\left(\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})}\right)\right] \mathrm{d}x\,\mathrm{d}y \\
&= -\iint p(\mathbf{x},\mathbf{y}) \ln p(\mathbf{x})\, \mathrm{d}x\,\mathrm{d}y + \iint p(\mathbf{x},\mathbf{y}) \ln p(\mathbf{x}|\mathbf{y})\, \mathrm{d}x\,\mathrm{d}y \\
&= -\int \left[\int p(\mathbf{x},\mathbf{y})\, \mathrm{d}y\right] \ln p(\mathbf{x})\, \mathrm{d}x + \mathbf{H}[\mathbf{x}|\mathbf{y}] \\
&= -\int p(\mathbf{x}) \ln p(\mathbf{x})\, \mathrm{d}x - \mathbf{H}[\mathbf{x}|\mathbf{y}] \\
&= \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}]
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{I}[\mathbf{x},\mathbf{y}] &= -\iint p(\mathbf{x},\mathbf{y}) \ln\left(\frac{p(\mathbf{y})}{p(\mathbf{x},\mathbf{y})/p(\mathbf{x})}\right) \mathrm{d}x\,\mathrm{d}y \\
&= -\iint p(\mathbf{x},\mathbf{y}) \left[\ln p(\mathbf{y}) - \ln\left(\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})}\right)\right] \mathrm{d}x\,\mathrm{d}y \\
&= -\iint p(\mathbf{x},\mathbf{y}) \ln p(\mathbf{y})\, \mathrm{d}x\,\mathrm{d}y + \iint p(\mathbf{x},\mathbf{y}) \ln p(\mathbf{y}|\mathbf{x})\, \mathrm{d}x\,\mathrm{d}y \\
&= -\int \left[\int p(\mathbf{x},\mathbf{y})\, \mathrm{d}x\right] \ln p(\mathbf{y})\, \mathrm{d}y + \mathbf{H}[\mathbf{y}|\mathbf{x}] \\
&= -\int p(\mathbf{y}) \ln p(\mathbf{y})\, \mathrm{d}y - \mathbf{H}[\mathbf{y}|\mathbf{x}] \\
&= \mathbf{H}[\mathbf{y}] - \mathbf{H}[\mathbf{y}|\mathbf{x}]
\end{aligned}
$$