Homework 1

Questions: https://github.com/SUSTech-ML-Course/Machine-Learning-Course/blob/05b4e0bd4063b9d486f34ce557720b519947cbe3/lesson-homework/hw1/HW1.md

Question 1

We have $Y = XW$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^M \\ 1 & x_2 & x_2^2 & \cdots & x_2^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^M \end{bmatrix} \quad W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{bmatrix}$$

Sum-of-squares error function

$$L = \frac{1}{2}(XW - Y)^2$$

We want to minimize it, so

$$\begin{aligned} \frac{\partial L}{\partial W} &= \frac{1}{2} \cdot \frac{\partial (XW-Y)^\top (XW-Y)}{\partial W} \\ &= \frac{1}{2} \cdot \frac{\partial \left(W^\top X^\top - Y^\top\right)(XW-Y)}{\partial W} \\ &= \frac{1}{2} \cdot \frac{\partial \left(W^\top X^\top XW - Y^\top XW - W^\top X^\top Y + Y^\top Y\right)}{\partial W} \\ &= \frac{1}{2} \cdot \left(X^\top XW + X^\top XW - X^\top Y - X^\top Y + 0\right) \\ &= X^\top XW - X^\top Y \\ &= 0 \end{aligned}$$

Finally we get $W = (X^\top X)^{-1} X^\top Y$

Question 2

(1)

$$\begin{aligned} p(apple) &= p(apple|r) \cdot p(r) + p(apple|b) \cdot p(b) + p(apple|g) \cdot p(g) \\ &= 0.3 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 \\ &= 0.34 \end{aligned}$$

(2)

$$
\begin{aligned}
p(g|orange) &= \frac{p(orange|g)\cdot p(g)}{p(orange)} \\
&= \frac{p(orange|g)\cdot p(g)}{p(orange|r)\cdot p(r)+p(orange|b)\cdot p(b)+p(orange|g)\cdot p(g)} \\
&= \frac{0.3\times0.6}{0.4\times0.2+0.5\times0.2+0.3\times0.6} \\
&= 0.5
\end{aligned}
$$

Question 3

$$
\begin{aligned}
E(X+Z) &= \iint (x+z)p(x,z)dxdz \\
&= \iint xp(x,z)dxdz + \iint zp(x,z)dxdz \\
&= \iint xp(x)p(z)dzdx + \iint zp(z)p(x)dxdz \\
&= \int xp(x)dx + \int zp(z)dz \\
&= E(X)+E(Z)
\end{aligned}
$$

$$
\begin{aligned}
var(X+Z) &= E((X+Z-E(X+Z))^2) \\
&= E((X+Z-E(X)-E(Z))^2) \\
&= E((X-E(X))^2+(Z-E(Z))^2-2(X-E(X))(Z-E(Z))) \\
&= E((X-E(X))^2+(Z-E(Z))^2-2Cov(X,Z)) \\
&= E((X-E(X))^2)+E((Z-E(Z))^2) \\
&= var(X)+var(Z)
\end{aligned}
$$

Question 4

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} P(X_i|\lambda) \\
lnL(\lambda) &= \sum_{i=1}^{n} lnP(X_i|\lambda)
\end{aligned}
$$

(1) If $X \sim Possion(\lambda)$, then $P(X|\lambda) = \frac{\lambda^X e^{-\lambda}}{X!}$

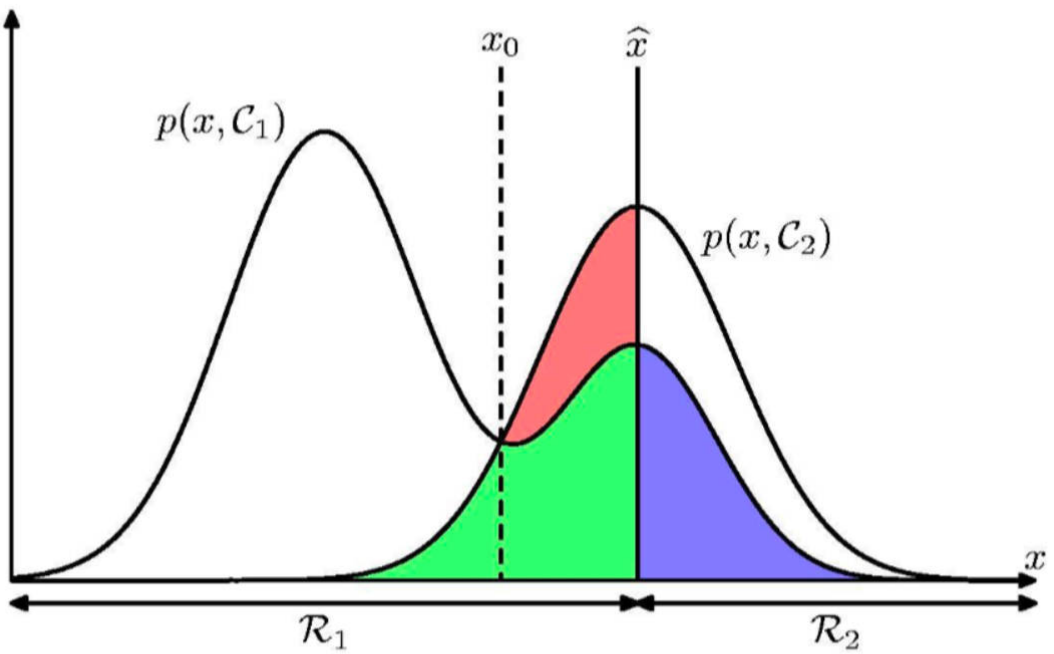
$$
\begin{aligned}
lnL(\lambda) &= \sum_{i=1}^{n} ln\frac{\lambda^{X_i}e^{-\lambda}}{X_i!} \\
&= \sum_{i=1}^{n}(X_i ln\lambda - \lambda - lnX_i!) \\
\frac{\partial lnL(\lambda)}{\partial\lambda} &= \sum_{i=1}^{n}(\frac{X_i}{\widehat{\lambda}}-1)=0 \\
\widehat{\lambda} &= \frac{\sum_{i=1}^{n} X_i}{n}
\end{aligned}
$$

(2) If $X \sim exp(\lambda)$, then $P(X|\lambda) = \frac{e^{-\frac{X}{\lambda}}}{\lambda}$

$$
\begin{aligned}
lnL(\lambda) &= \sum_{i=1}^{n} ln \frac{e^{-\frac{X_i}{\lambda}}}{\lambda} \\
&= \sum_{i=1}^{n} (-\frac{X_i}{\lambda} - ln\lambda) \\
\frac{\partial lnL(\lambda)}{\partial \lambda} &= \sum_{i=1}^{n} (\frac{X_i}{\widehat{\lambda}^2} - \frac{1}{\widehat{\lambda}}) \\
&= \frac{\sum_{i=1}^{n} X_i}{\widehat{\lambda}^2} - \frac{n}{\widehat{\lambda}} = 0 \\
\widehat{\lambda} &= \frac{\sum_{i=1}^{n} X_i}{n}
\end{aligned}
$$

Question 5

(a)



$$
p(correct) = \int_0^{\widehat{x}} p(x, C_1) + \int_{\widehat{x}}^{\infty} p(x, C_2)
$$

$$
p(mistake) = \int_0^{\widehat{x}} p(x, C_2) + \int_{\widehat{x}}^{\infty} p(x, C_1)
$$

(b)

$$
\begin{aligned}
E[L(\mathbf{t},\mathbf{y}(\mathbf{x}))] &= \iint \|\mathbf{y}(\mathbf{x})-\mathbf{t}\|^2 p(\mathbf{x},\mathbf{t})d\mathbf{x}d\mathbf{t} \\
\frac{\partial E}{\partial \mathbf{y}(\mathbf{x})} &= \int 2(\mathbf{y}(\mathbf{x})-\mathbf{t})p(\mathbf{x},\mathbf{t})d\mathbf{t} = 0 \\
\int \mathbf{y}(\mathbf{x})p(\mathbf{x},\mathbf{t})d\mathbf{t} &= \int \mathbf{t}p(\mathbf{x},\mathbf{t})d\mathbf{t} \\
\mathbf{y}(\mathbf{x})p(\mathbf{x}) &= \int \mathbf{t}p(\mathbf{x},\mathbf{t})d\mathbf{t} \\
\mathbf{y}(\mathbf{x}) &= \frac{\int \mathbf{t}p(\mathbf{x},\mathbf{t})d\mathbf{t}}{p(\mathbf{x})} \\
&= \int \mathbf{t}p(\mathbf{t}|\mathbf{x})d\mathbf{t} \\
&= E_{\mathbf{t}}[\mathbf{t}|\mathbf{x}]
\end{aligned}
$$

Question 6

(a)

$$
\begin{aligned}
\mathbf{H}[\mathbf{x}] &= -\int_{-\infty}^{\infty} p(\mathbf{x})\ln p(\mathbf{x})d\mathbf{x} \\
&= -\int_{-\infty}^{\infty} \frac{e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} ln \frac{e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} d\mathbf{x} \\
&= -\int_{-\infty}^{\infty} \frac{e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} ln(\frac{1}{\sqrt{2\pi}\sigma} - \frac{(\mathbf{x}-\mu)^2}{2\sigma^2})d\mathbf{x} \\
&= \frac{1}{\sqrt{2\pi}\sigma}(\int_{-\infty}^{\infty} ln(\sqrt{2\pi}\sigma)e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}d\mathbf{x} + \int_{-\infty}^{\infty} \frac{(\mathbf{x}-\mu)^2}{2\sigma^2}e^{-\frac{(\mathbf{x}-\mu)^2}{2\sigma^2}}d\mathbf{x}) \\
&= ln\sqrt{2\pi}\sigma \int_{-\infty}^{\infty} p(\mathbf{x})d\mathbf{x} + \frac{\sqrt{2}\sigma}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \mathbf{y}^2 e^{-\mathbf{y}^2}d\mathbf{y} \\
&= ln\sqrt{2\pi}\sigma + \frac{1}{\sqrt{\pi}}(-\frac{1}{2}\mathbf{y}e^{-\mathbf{y}^2}|_{-\infty}^{\infty} + \frac{1}{2}\int_{-\infty}^{\infty} e^{-\mathbf{y}^2}d\mathbf{y}) \\
&= ln\sqrt{2\pi}\sigma + \frac{1}{\sqrt{\pi}}(0 + \frac{\sqrt{\pi}}{2}) \\
&= ln\sqrt{2\pi}\sigma + \frac{1}{2}
\end{aligned}
$$

(b)

(1) X, Y are both discrete

$$
\begin{aligned}
\mathbf{I}(\mathbf{x},\mathbf{y}) &= \sum_{\mathbf{y}}\sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \\
&= \sum_{\mathbf{y}}\sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} - \sum_{\mathbf{y}}\sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})ln\frac{1}{p(\mathbf{x})} \\
&= \sum_{\mathbf{y}}\sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})lnp(\mathbf{x}) + \sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} \\
&= \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}]
\end{aligned}
$$

And we can derive $\mathbf{I}(\mathbf{x},\mathbf{y}) = \mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{y}|\mathbf{x})$ in the same way

(2) X, Y are both continous

$$
\begin{aligned}
\mathbf{I}(\mathbf{x},\mathbf{y}) &= \int_{\mathbf{y}}\int_{\mathbf{x}} p(\mathbf{x},\mathbf{y}) ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} d\mathbf{x}d\mathbf{y} \\
&= \int_{\mathbf{y}}\int_{\mathbf{x}} p(\mathbf{x},\mathbf{y}) ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} d\mathbf{x}d\mathbf{y} - \int_{\mathbf{y}}\int_{\mathbf{x}} p(\mathbf{x},\mathbf{y}) ln\frac{1}{p(\mathbf{x})} d\mathbf{x}d\mathbf{y} \\
&= \int_{\mathbf{y}}\int_{\mathbf{x}} p(\mathbf{x},\mathbf{y}) ln p(\mathbf{x}) d\mathbf{x}d\mathbf{y} + \int_{\mathbf{x}} p(\mathbf{x},\mathbf{y}) ln\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})} d\mathbf{x} \\
&= \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}]
\end{aligned}
$$

And we can derive $\mathbf{I}(\mathbf{x},\mathbf{y}) = \mathbf{H}(\mathbf{y}) - \mathbf{H}(\mathbf{y}|\mathbf{x})$ in the same way