CS405 Homework #1

SID: 12110524

Name: 陈康睿

Question 1

Sum-of-squares error $E(w)$ is defined as:

$$
\begin{aligned}
E(w) &= \sum_{i=1}^{N}[y(x_i, w) - t_i]^2 \\
&= \sum_{i=1}^{N}[\sum_{j=0}^{M} w_j x_i^j - t_i]^2
\end{aligned}
$$

where N is the total number of sample points, (x_j, t_j) is a pair of input and target output.

Derivatives with respect to w_k is:

$$
\frac{\partial E}{\partial w_k} = 2\sum_{i=1}^{N} x_i^k (\sum_{j=0}^{M} w_j x_i^j - t_i)
$$

To minimize $E(w)$, set the derivatives equal to 0, then the equation can be transferred to:

$$
\sum_{i=1}^{N} x_i^k \sum_{j=0}^{M} w_j x_i^j = \sum_{i=1}^{N} x_i^k t_i
$$

which can be expressed in matrix form:

$$
\begin{bmatrix} x_1^k & \cdots & x_N^k \end{bmatrix}
\begin{bmatrix} x_1^0 & \cdots & x_1^M \\ \vdots & \ddots & \vdots \\ x_N^0 & \cdots & x_N^M \end{bmatrix}
\begin{bmatrix} w_0 \\ \vdots \\ w_M \end{bmatrix}
=
\begin{bmatrix} x_1^k & \cdots & x_N^k \end{bmatrix}
\begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}
$$

Then consider all equations:

$$
\begin{bmatrix} x_1^0 & \cdots & x_N^0 \\ \vdots & \ddots & \vdots \\ x_1^M & \cdots & x_N^M \end{bmatrix}
\begin{bmatrix} x_1^0 & \cdots & x_1^M \\ \vdots & \ddots & \vdots \\ x_N^0 & \cdots & x_N^M \end{bmatrix}
\begin{bmatrix} w_0 \\ \vdots \\ w_M \end{bmatrix}
=
\begin{bmatrix} x_1^0 & \cdots & x_N^0 \\ \vdots & \ddots & \vdots \\ x_1^M & \cdots & x_N^M \end{bmatrix}
\begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix}
$$

Let

$$X = \begin{bmatrix} x_1^0 & \cdots & x_1^M \\ \vdots & \ddots & \vdots \\ x_N^0 & \cdots & x_N^M \end{bmatrix}$$
$$A = X^T X$$
$$w = \begin{bmatrix} w_0 \\ \vdots \\ w_M \end{bmatrix}$$
$$t = \begin{bmatrix} t_0 \\ \vdots \\ t_N \end{bmatrix}$$
$$b = X^T t$$

now the above linear equation can be represented by:

$$Aw = b$$

So the solution should be (according to least-square inverse):

$$w = (A^T A)^{-1} A^T b = (X^T X)^{-1} X^T t$$

Question 2

(a)

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

(b)

$$\begin{aligned} p(g|orange) &= \frac{p(orange|g) \times p(g)}{p(orange)} \\ &= \frac{0.3 \times 0.6}{0.4 \times 0.2 + 0.5 \times 0.2 + 0.4 \times 0.6} \\ &= 0.5 \end{aligned}$$

Question 3

(a)

Continuous

$$\begin{aligned} \mathbb{E}(X+Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+z) p(x,z) dx dz \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x p(x,z) dz dx + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z p(x,z) dx dz \\ &= \int_{-\infty}^{\infty} x P_X(x) dx + \int_{-\infty}^{\infty} z P_Z(z) dz \\ &= \mathbb{E}(x) + \mathbb{E}(Z) \end{aligned}$$

Discrete

$$\begin{aligned}\mathbb{E}(X+Z) &= \sum_x \sum_z (x+z)p(x,z) \\ &= \sum_x \sum_z xp(x,z) + \sum_z \sum_x zp(x,z) \\ &= \sum_x xP_X(x) + \sum_z zP_Z(z) \\ &= \mathbb{E}(x) + \mathbb{E}(Z)\end{aligned}$$

(b)

$$\begin{aligned}var(X+Z) &= \mathbb{E}[(X+Z)^2] - \mathbb{E}(X+Z)^2 \\ &= \mathbb{E}(X^2+2XZ+Z^2) - [\mathbb{E}(X)^2 + 2\mathbb{E}(X)\mathbb{E}(Z) + \mathbb{E}(Z)^2] \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 + \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 + 2\mathbb{E}(XZ) - 2\mathbb{E}(X)\mathbb{E}(Z) \\ &= var(X) + var(Z)\end{aligned}$$

Question 4

(a)

$$L(\lambda) = \prod_{i=1}^{n} P(X_i|\lambda)$$

$$\implies \ln L(\lambda) = \ln(\prod_{i=1}^{n} \frac{1}{X_i!}) + \ln\lambda \sum_{i=1}^{n} X_i - n\lambda$$

$$\implies \frac{d\ln L(\lambda)}{d\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} X_i - n$$

$$Let\ \frac{d\ln L(\lambda)}{d\lambda} = 0$$

$$Then\ the\ MSE\ is\ \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

(b)

$$L(\lambda) = \prod_{i=1}^{n} P(X_i|\lambda)$$

$$\implies \ln L(\lambda) = -n\ln\lambda - \frac{1}{\lambda}\sum_{i=1}^{n} X_i$$

$$\implies \frac{d\ln L(\lambda)}{d\lambda} = \frac{1}{\lambda^2}\sum_{i=1}^{n} X_i - \frac{n}{\lambda}$$

$$Let\ \frac{d\ln L(\lambda)}{d\lambda} = 0$$

$$Then\ the\ MSE\ is\ \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Question 5

(a)

$$p(correct) = p(x \in \mathcal{R}_1, \mathcal{C}_1) + p(x \in \mathcal{R}_2, \mathcal{C}_2) = \int_{R_1} p(x, \mathcal{C}_1)dx + \int_{R_2} p(x, \mathcal{C}_2)dx$$

$$p(mistake) = p(x \in \mathcal{R}_1, \mathcal{C}_2) + p(x \in \mathcal{R}_2, \mathcal{C}_1) = \int_{R_1} p(x, \mathcal{C}_2)dx + \int_{R_2} p(x, \mathcal{C}_1)dx$$

(b)

As multi-dimension error can be seen as the sum of error in each dimension, for each dimension i :

$$\frac{\partial \mathbb{E}[L(t, y_i(x_I))]}{\partial y(x)} = 2\int \{y_i(x_I) - t\}p(x,t)dt = 0$$

$$\Rightarrow y_i(x_i) = \frac{\int t_i p(x_i, t_i)dt}{p(x_i)} = \mathbb{E}_{t_i}[t_i|x_i]$$

Then splice them: $y(x) = \mathbb{E}[t|x]$

Question 6

(a)

$$\begin{aligned}
H(X) &= -\int \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}})dx \\
&= -\int \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln(\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}})dx \ \ (let\ t = \frac{x-\mu}{\sqrt{2}\sigma},\ then\ \frac{dt}{dx} = \frac{1}{\sqrt{2}\sigma}) \\
&= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-t^2}[t^2 + \ln(\sqrt{2\pi}\sigma)]dx \\
&= \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-t^2}dt + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} t^2 e^{-t^2}dt \\
&= \ln(\sqrt{2\pi}\sigma) + \frac{1}{2}
\end{aligned}$$

(b)

Continus Distribution

$$\begin{aligned}
\mathbf{I}[\mathbf{X}, \mathbf{Y}] &= \int_{\mathcal{X}}\int_{\mathcal{Y}} \ln(\frac{P_{X,Y}(x,y)}{P_X(x)P_Y(y)})dydx \\
&= \int_{\mathcal{X}}\int_{\mathcal{Y}} P_{X,Y}(x,y)\ln P(x|y)dydx - \int_{\mathcal{X}} P_X(x)\ln(P_X(x))dx \\
&= \mathbf{H}[\mathbf{X}] - \mathbf{H}[\mathbf{X}|\mathbf{Y}]
\end{aligned}$$

Similarly, $\mathbf{H}[\mathbf{Y}] - \mathbf{H}[\mathbf{Y}|\mathbf{X}]$.

Discrete Distribution

Similar to continus distribution by replacing integration with summation.