Question 1: $y(x,w) = w_0 + w_1 x + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j$

$L(w) = \frac{1}{2}\sum_{i=1}^{N}\left[y_i - \sum_{j=0}^{M} w_j x_i^j\right]^2$ 分别对 $w_1, w_2 \cdots w_M$ 求偏导

有 $\frac{\partial L}{\partial (w_k)} = \sum_{i=1}^{n}\left[y - \sum_{j=0}^{M} w_j x^j\right] x^k$ 分别令 $\frac{\partial L}{\partial (w_k)} = 0$ 得

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i & \cdots & \sum_{i=1}^{n} x_i^M \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \cdots & \sum_{i=1}^{n} x_i^{M+1} \\ \vdots & & & \vdots \\ \sum_{i=1}^{n} x_i^M & \sum_{i=1}^{n} x_i^{M+1} & \cdots & \sum_{i=1}^{n} x_i^{2M} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_M \end{bmatrix} = \cancel{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}} \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \\ \vdots \\ \sum_{i=1}^{n} x_i^k y_i \end{bmatrix}$$

有 $\begin{bmatrix} 1 & x_1 & \cdots & x_1^M \\ 1 & x_2 & \cdots & x_2^M \\ \vdots & & & \\ 1 & x_n & \cdots & x_n^M \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $w = (X^T X)^{-1} X^T y$

Question 2:

$P(apple) = P(r)\cdot\frac{3}{3+4+3} + P(b)\cdot\frac{1}{1+1} + P(g)\cdot\frac{3}{3+4+3}$

$= 0.34$

$P(g|orange) = \frac{P(g)\cdot\frac{3}{3+4+3}}{P(orange)} = \frac{0.6\cdot\frac{3}{10}}{P(r)\cdot\frac{4}{3+4+3} + P(b)\cdot\frac{1}{1+1} + P(g)\frac{3}{3+4+3}}$

~~Questi~~

$= \frac{\frac{1.8}{10}}{0.36} = \cancel{0.72}\ 0.5$

Question 3:

$E(x+z) = \frac{1}{n}\sum_{i=1}^{n}(x_i+z_i) = \frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{n}\sum_{i=1}^{n} z_i$

$= E(x) + E(z)$

$\because$ x and y are independent variables

$var(x+z) = E(x+z)^2 - [E(x+z)]^2$

$= E(x^2 + 2xz + z^2) - (Ex + Ez)^2$

$= E(x^2) + E(z^2) + 2E(xz) - (Ex)^2 - 2ExEz - E(z)^2$

$= var(x) + var(z) + 2E(xz) - 2ExEz$

$= var(x) + var(z)$

Question 4:

$P(X|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$

$L(x_1, x_2 \cdots, x_n|\lambda) = \prod_{i=1}^{n}\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = e^{-n\lambda}\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}$

$\ln L = -n\lambda + \sum_{i=1}^{n}(x_i \ln\lambda - \ln x_i)$

$\frac{d\ln L}{d\lambda} = -n + \sum_{i=1}^{n}\frac{x_i}{\lambda}$

令 $\frac{d\ln L}{d\lambda} = 0$ 得 $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

exponential distribution:

$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}$

$L(x_1, x_2, \cdots, x_n|\lambda) = \prod_{i=1}^{n}\frac{1}{\lambda} e^{-\frac{x_i}{\lambda}} = \frac{1}{\lambda^n}\cdot e^{-\frac{\sum_{i=1}^{n} x_i}{\lambda}}$

$\ln L = -n\ln\lambda + \left(-\frac{\sum_{i=1}^{n} x_i}{\lambda}\right)$

$\frac{d\ln L}{d\lambda} = -\frac{n}{\lambda} + \frac{\sum_{i=1}^{n} x_i}{\lambda^2}$ 令 $\frac{d\ln L}{d\lambda} = 0$

得 $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Question 5

a) $p(mistake) = \int_{R_1} p(x, c_2)\,dx + \int_{R_2} p(x, c_1)\,dx$

$p(correct) = \int_{R_1} p(x, c_1)\,dx + \int_{R_2} p(x, c_2)\,dx$

b) $E[L(\vec{t}, \vec{y}(x))] = \iint \|\vec{y}(x) - \vec{t}\|^2 p(\vec{x}, \vec{t})\,dx\,dt$

$\|\vec{y}(x) - \vec{t}\|^2 = \|\vec{y}(x) - \vec{E}_t[t|x] + \vec{E}_t[t|x] - \vec{t}\|^2$

$= \|\vec{y}(x) - \vec{E}_t[t|x]\|^2 +$

Question 6:

a) $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$H[x] = -\int p(x)\ln p(x)\,dx$

$H[x] = -\int \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$

$= -\frac{1}{\sqrt{2\pi}\sigma}\int e^{-\frac{(x-\mu)^2}{2\sigma^2}}\left(-\ln(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}\right)dx$

$= \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{2\pi}\sigma}\int e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx - \frac{1}{\sqrt{2\pi}\sigma}\int e^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot\frac{(x-\mu)^2}{2\sigma^2}\,dx$

$= \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{2\pi}\sigma}\sqrt{2}\sigma\int e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2} d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$

$+ \frac{1}{\sqrt{2\pi}\sigma}\sqrt{2}\sigma\int e^{-\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2}\cdot\frac{(x-\mu)^2}{2\sigma^2}\,d\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)$

记 $y = \frac{x-\mu}{\sqrt{2}\sigma}$

有 $H[x] = \frac{\ln(\sqrt{2\pi}\sigma)}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2} dy + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2} y^2 dy$

$$= \ln(\sqrt{2\pi}\sigma) + \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-y^2} y^2 dy$$

$$= \ln(\sqrt{2\pi}\sigma) + \frac{1}{\sqrt{\pi}} \cdot \left(-\frac{1}{2}\right) \cdot \left(-\int_{-\infty}^{+\infty} e^{-y^2} dy\right)$$

$$= \ln(\sqrt{2\pi}\sigma) + \frac{1}{2}$$

$$= \frac{1}{2}\{1 + \ln(2\pi\sigma^2)\}$$

b)

$$\cancel{KL(p\|q) = -\int p(x) \ln\left\{\frac{q(x)}{p(x)}\right\} dx}$$

$$\cancel{I[x,y] = KL(p(x,y)\|p(x)p(y))}$$

$$\cancel{= -\iint p(x,y) \cdot \ln\left(\frac{p(x)p(y)}{p(x,y)}\right) dxdy}$$

$$\cancel{I[x,y] = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}}$$

$$\cancel{= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}}$$

b)

$$I[x,y] = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)} - \sum_{x,y} p(x,y) \ln p(y)$$

$$= \sum_{x,y} p(x)p(y|x) \ln p(y|x) - \sum_{x,y} p(x,y) \ln p(y)$$

$$= \sum_{x} p(x)\left(\sum_{y} p(y|x) \ln p(y|x)\right) - \sum_{y} \ln p(y)\left(\sum_{x} p(x,y)\right)$$

$$= -\sum_{x} p(x) H(Y|X=x) - \sum_{y} \ln p(y) p(y)$$

$$= H(Y) - H(Y|X)$$

$$I[x,y] = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(y)} - \sum_{x,y} p(x,y) \ln p(x)$$

$$= \sum_{x,y} p(x,y) \ln p(x|y) - \sum_{x,y} p(x,y) \ln p(x)$$

$$= \sum_{x,y} p(y) \cdot p(x|y) \cdot \ln(x|y) - \sum_{x} \ln p(x) \left(\sum_{y} p(x,y)\right)$$

$$= -\sum_{y} p(y) H(x|Y=y) - \sum_{x} p(x) \cdot \ln p(x)$$

$$= H[x] - H[x|y]$$

$$\cancel{I[x,y] = H[x] - H[Y|X] = H[x]}$$

$$I[x,y] = H[Y] - H[Y|X] = H[X] - H[X|y]$$