Q1.

$$y(x, w) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{i=0}^{M} w_i x^i$$

rewrite : $Y = XW$

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

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

$$\frac{\partial L}{\partial W} = \frac{1}{2} \cdot \frac{\partial (XW-Y)^T(XW-Y)}{\partial W}$$

$$= \frac{1}{2} \cdot \frac{\partial (W^TX^T - Y^T)(XW-Y)}{\partial W}$$

$$= \frac{1}{2} \cdot \frac{\partial (W^TX^TXW - Y^TXW - W^TX^TY + Y^TY)}{\partial W}$$

$$= \frac{1}{2} \cdot (X^TXW + X^TXW - X^TY - X^TY + 0)$$

$$= X^TXW - X^TY$$

$$= 0$$

$$\Rightarrow X^TXW = X^TY$$

$$\Rightarrow W = (X^TX)^{-1} X^TY$$

Q2.

$p(apple \mid r) = \frac{3}{10}$ $p(orange \mid r) = \frac{4}{10}$ $p(lime \mid r) = \frac{3}{10}$

$p(apple \mid b) = \frac{1}{2}$ $p(orange \mid b) = \frac{1}{2}$ $p(lime \mid b) = 0$

$p(apple \mid g) = \frac{3}{10}$ $p(orange \mid g) = \frac{3}{10}$ $p(lime \mid g) = \frac{4}{10}$

① $p(apple) = p(apple \mid r) \cdot p(r) + p(apple \mid b) \cdot p(b) + p(apple \mid g) \cdot p(g)$

$= \frac{3}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6$

$= 0.34$

② $p(g \mid orange) = \frac{p(orange \mid g) \cdot p(g)}{p(orange)} = \frac{\frac{3}{10} \times 0.6}{\frac{4}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6} = 0.5$

Q3.

For $E[X+Z] = E(X) + E(Z)$

① when X. Z is discrete:

$$E(X+Z) = \sum_i \sum_j (X_i + Z_j) \cdot P(X_i, Z_j)$$
$$= \sum_i \sum_j X_i \cdot P(X_i, Z_j) + \sum_i \sum_j Z_j P(X_i, Z_j)$$
$$= \sum_i \sum_j X_i P(X_i) \cdot P(Z_j) + \sum_i \sum_j Z_j P(X_i) \cdot P(Z_j)$$
$$= \sum_i X_i P(X_i) + \sum_j Z_j P(Z_j)$$
$$= E(X) + E(Z)$$

② when X is continuous:

$$E(X+Z) = \iint (x+z) p(x,z) \cdot dx dz$$
$$= \iint x \cdot p(x,z) \cdot dx dz + \iint z p(x,z) dx dz$$
$$= \iint x \cdot p(x) \cdot p(z) \cdot dz dx + \iint z \cdot p(z) \cdot p(x) \cdot dx dz$$
$$= \int x p(x) dx + \int z p(z) dz$$
$$= E(X) + E(Z)$$

For $var[X+Z] = var[X] + var[Z]$

$$var[X+Z] = E[((X+Z) - E[X+Z])^2]$$
$$= E[(X+Z - E(X) - E(Z))^2]$$
$$= E[(X - E(X))^2 + (Z - E(Z))^2 - 2(X - E(X)) \cdot (Z - E(Z))]$$
$$= E[(X - E(X))^2 + (Z - E(Z))^2 - 2 Cov(X, Z)]$$
$$= E[(X - E(X))^2] + E[(Z - E(Z))^2]$$
$$= var[X] + var[Z]$$

Q4.

For $X \sim Possion(\lambda)$, $\quad L(\lambda) = \prod_{i=1}^{n} P(X_i | \lambda)$

$$\Rightarrow \ln L(\lambda) = \sum_{i=1}^{n} \ln \frac{\lambda^{X_i} \cdot e^{-\lambda}}{X!} = \sum_{i=1}^{n} (X_i \cdot \ln\lambda + (-\lambda) - \ln(X!))$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \sum_{i=1}^{n}\left(\frac{X_i}{\lambda} - 1\right) = 0$$

$$\Rightarrow \quad \lambda = \frac{1}{n}\cdot \sum_{i=1}^{n} X_i$$

For $X \sim \exp(\lambda)$, $\quad L(\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda}\cdot e^{-\frac{X_i}{\lambda}}$

$$\Rightarrow \quad \ln L(\lambda) = \sum_{i=1}^{n} \ln\left(\frac{1}{\lambda}\cdot e^{-\frac{X_i}{\lambda}}\right) = \sum_{i=1}^{n}\left(-\frac{X_i}{\lambda} - \ln\lambda\right)$$

$$\frac{\partial \ln L(\lambda)}{\partial \lambda} = \sum_{i=1}^{n}\left(\frac{X_i}{\lambda^2} - \frac{1}{\lambda}\right) = 0$$

$$\Rightarrow \quad \lambda = \frac{1}{n}\cdot \sum_{i=1}^{n} X_i$$

Q5

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

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

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

$$\frac{\partial E}{\partial y(x)} = \int 2(y(x) - t)\cdot p(x, t)\, dt = 0.$$

$$\Rightarrow \quad \int y(x)\cdot p(x, t)\, dt = \int t\, p(x, t)\, dt$$

$$\Rightarrow \quad y(x)\cdot p(x) = \int t\cdot p(x, t)\, dt$$

$$\Rightarrow \quad y(x) = \frac{\int t\cdot p(x, t)\, dt}{p(x)} = \int t\cdot p(t|x)\cdot dt = E(t|x)$$

Q6

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

$$\Rightarrow H[X] = -\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot\left(\ln\frac{1}{\sqrt{2\pi}\sigma} - \frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

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

$$+ \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \frac{(x-\mu)^2}{2\sigma^2}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

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

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

$$= \ln(\sqrt{2\pi}\sigma) + \frac{1}{\sqrt{\pi}}\cdot\left(0 + \frac{\sqrt{\pi}}{2}\right)$$

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

(b) $I[X,Y] = KL(P(X,Y) \,||\, P(X)\cdot P(Y))$

① When X, Y are both discrete:

$$I[X,Y] = \sum_y \sum_x p(x,y)\cdot \ln \frac{p(x,y)}{p(x)\cdot p(y)}$$

$$= \sum_y \sum_x p(x,y)\cdot \ln \frac{p(x,y)}{p(y)} - \sum_y \sum_x p(x,y)\cdot \ln \frac{1}{p(x)}$$

$$= \sum_y \sum_x p(x,y)\cdot \ln p(x|y) + \sum_x \cdot p(x)\cdot \ln p(x)$$

$$= -H[X|Y] + H[X]$$

$$= H[X] - H[X|Y]$$

In the same way, we derive $I[X,Y] = H[Y] - H[Y|X]$

② When X, Y are both continuous:

$$I[X,Y] = \int_y \int_x p(x,y)\cdot \ln \frac{p(x,y)}{p(x)\cdot p(y)}\, dx dy$$

$$= \int_y \int_x p(x,y)\cdot \ln \frac{p(x,y)}{p(y)}\, dx dy - \int_y \int_x p(x,y)\cdot \ln \frac{1}{p(x)}\, dx dy$$

$$= \int_y \int_x p(x,y)\cdot \ln p(x|y)\, dx dy + \int_x p(x)\cdot \ln p(x)\, dx$$

$$= -H[X|Y] + H[X]$$

$$= H[X] - H[X|Y]$$

In the same way, we derive $I[X,Y] = H[Y] - H[Y|X]$