Question 1

Minimize

$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\Big(y(x_n, \mathbf{w}) - t_n\Big)^2$$

where

$$\begin{aligned}\frac{\partial E(w)}{w_j} &= \sum_{i=1}^{N}\Big(y(x_i, \mathbf{w}) - t_i\Big)\frac{\partial y(x_i, \mathbf{w})}{\partial w_j} \\ &= \sum_{i=1}^{N}\Big(y(x_i, \mathbf{w}) - t_i\Big)x_i^j\end{aligned}$$

Next, set the derivatives equal to zero to find the coefficients that minimize the error:

$$\sum_{i=1}^{N}\Big(y(x_i, \mathbf{w}) - t_i\Big)x_i^j = 0$$

Now, we can solve this equation for each coefficient w_i separately. This will give a system of equations, one for each coefficient w_i.

Question 2

$$\begin{aligned}P(\text{apple}) &= P(\text{apple} \mid \text{r})P(\text{r}) + P(\text{apple} \mid \text{b})P(\text{b}) + P(\text{apple} \mid \text{g})P(\text{g}) \\ &= \frac{3}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 \\ &= 0.34\end{aligned}$$

$$\begin{aligned}P(\text{orange}) &= P(\text{orange} \mid \text{r})P(\text{r}) + P(\text{orange} \mid \text{b})P(\text{b}) + P(\text{orange} \mid \text{g})P(\text{g}) \\ &= \frac{4}{10} \times 0.2 + \frac{1}{2} \times 0.2 + \frac{3}{10} \times 0.6 \\ &= 0.36\end{aligned}$$

$$\begin{aligned}P(\text{g} \mid \text{orange}) &= \frac{P(\text{orange} \mid \text{g})P(\text{g})}{P(\text{orange})} \\ &= \frac{\frac{3}{10} \times 0.6}{0.36} \\ &= 0.5\end{aligned}$$

Question 3

$$\begin{aligned}\mathbb{E}[X+Z] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+z)f(x,z)dzdx \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf(x,z)dzdx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}zf(x,z)dzdx \\ &= \mathbb{E}[X] + \mathbb{E}[Z]\end{aligned}$$

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

Question 4

(1)

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} P(X=X_i|\lambda) \\
&= \prod_{i=1}^{n} \frac{\lambda^{X_i}e^{-\lambda}}{X_i!}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{d\ln L(\lambda)}{d\lambda} &= \frac{\sum_{i=1}^{n} X_i}{\lambda} - n \\
&= 0
\end{aligned}
$$

It is solved that

$$
\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i
$$

(2)

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} f(x=X_i|\lambda) \\
&= \prod_{i=1}^{n} \frac{1}{\lambda}e^{-\frac{X_i}{\lambda}}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{d\ln L(\lambda)}{d\lambda} &= \frac{\sum_{i=1}^{n} X_i}{\lambda^2} - \frac{n}{\lambda} \\
&= 0
\end{aligned}
$$

It is solved that

$$
\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i
$$

Question 5

(a)

$$
\begin{aligned}
p(\text{mistake}) &= p(x \in \mathcal{R}_1, \mathcal{C}_2) + p(x \in \mathcal{R}_2, \mathcal{C}_1) \\
&= \int_{\mathcal{R}_1} p(x, \mathcal{C}_2) dx + \int_{\mathcal{R}_2} p(x, \mathcal{C}_1) dx
\end{aligned}
$$

$$
p(\text{correct}) = 1 - p(\text{mistake})
$$

(b)

$$
\begin{aligned}
\frac{\delta \mathbb{E}[L(t, y(\mathbf{x}))]}{\delta y(\mathbf{x})} &= 2 \int \{y(\mathbf{x}) - t\} p(\mathbf{x}, t) dt \\
&= 0
\end{aligned}
$$

求解$y(\mathbf{x})$，使用概率的加和规则和乘积规则，我们得到

$$
\begin{aligned}
y(\mathbf{x}) &= \frac{\int tp(\mathbf{x}, t) dt}{p(\mathbf{x})} \\
&= \int tp(t|\mathbf{x}) dt \\
&= \mathbb{E}_t[t|\mathbf{x}]
\end{aligned}
$$

Question 6

(a)

$$
\begin{aligned}
\mathbf{H}[\mathbf{X}] &= -\int p(x) \ln p(x) dx \\
&= -\int p(x) \ln \Big(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Big) dx \\
&= \int p(x) \Big(\ln \sqrt{2\pi}\sigma + \frac{(x-\mu)^2}{2\sigma^2} \Big) dx \\
&= \ln \sqrt{2\pi}\sigma \int p(x) dx + \frac{1}{2\sigma^2} \int (x-\mu)^2 p(x) dx \\
&= \ln \sqrt{2\pi}\sigma + \frac{1}{2} \\
&= \frac{1}{2} \{\ln(2\pi\sigma^2) + 1\}
\end{aligned}
$$

(b)

$$
\begin{aligned}
I[\boldsymbol{y}, \boldsymbol{x}] &\equiv \mathrm{KL}(p(\boldsymbol{y}, \boldsymbol{x}) || p(\boldsymbol{y}) p(\boldsymbol{x})) \\
&= -\int \int p(\boldsymbol{y}, \boldsymbol{x}) \ln \Big(\frac{p(\boldsymbol{y}) p(\boldsymbol{x})}{p(\boldsymbol{y}, \boldsymbol{x})} \Big) d\boldsymbol{x} d\boldsymbol{y}
\end{aligned}
$$

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

Similarity, we derive

$$I[\boldsymbol{x}, \boldsymbol{y}] = H[\boldsymbol{y}] - H[\boldsymbol{y}|\boldsymbol{x}]$$

Hence,

$$\mathbf{I}[\mathbf{x}, \mathbf{y}] = \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}] = \mathbf{H}[\mathbf{y}] - \mathbf{H}[\mathbf{y}|\mathbf{x}]$$