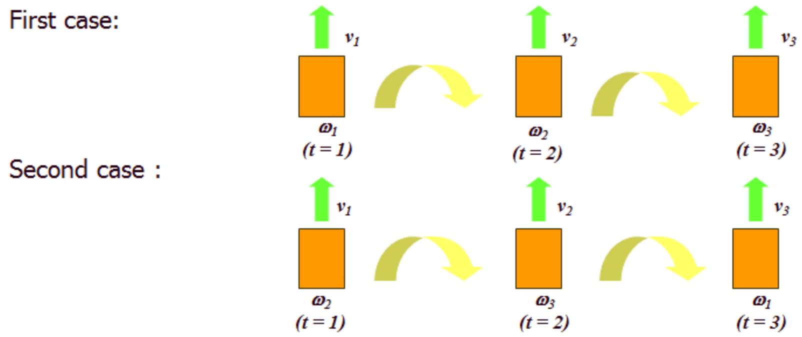
$$p(\{v_1, v_2, v_3\}) = p(w_1)\, p(v_1|w_1)\, p(w_2|w_1)\, p(v_2|w_2)\, p(w_3|w_2)\, p(v_3|w_3) + \cdots$$

Question Three
Answer:



$$P(\{v_1, v_2, v_3\}) = P(\omega_1)P(v_1 | \omega_1)P(\omega_2 | \omega_1)P(v_2 | \omega_2)P(\omega_3 | \omega_2)P(v_3 | \omega_3)$$
$$+ P(\omega_2)P(v_1 | \omega_2)P(\omega_3 | \omega_2)P(v_2 | \omega_3)P(\omega_1 | \omega_3)P(v_3 | \omega_1)$$

Question Four
Answer:
$z = X\alpha$

$\alpha = (X^T X + \gamma I)^{-1} X^T z$ (if $X^T X$ is singular, otherwise $\alpha = X^T X^{-1} X^T z$),

where γ is a small positive constant and I is the identity matrix.

We exploit construction residual of each class to classify the test sample. After representation coefficient α is obtained, we use training samples of each class to represent the test sample as

$$g_i = X_i \alpha_i,$$

where X_i denotes all training samples of ith class and α_i is their corresponding representation coefficients.

Then the construction residual of each class is calculated as follows $r_i = \|z - g_i\|_2$.

If $k = \arg\min_i r_i$, then the test sample will be classified to the kth class.

Question Five
Answer:

$\theta_1 = u$, $\theta_2 = \sigma^2$.

$$p(D|\theta)=\prod_{k=1}^{n} p(x_k|\theta),\quad \nabla_\theta=\begin{bmatrix}\frac{\partial}{\partial\theta_1}\\ \frac{\partial}{\partial\theta_2}\end{bmatrix},$$

$$l(\theta)=\ln p(D|\theta)\quad,\quad \hat{\theta}=\arg\max_\theta l(\theta),$$

$$l(\theta)=\sum_{k=1}^{n}\ln p(x_k|\theta)$$

$$p(x_k|\theta)=\frac{1}{\sqrt{2\pi\theta_2}}\exp\left(-\frac{(x_k-\theta_1)^2}{2\theta_2}\right)$$

$$\nabla_\theta l=\sum_{k=1}^{n}\nabla_\theta \ln p(x_k|\theta)=0$$

$$l(\theta)=\sum_{k=1}^{n}\ln p(x_k|\theta)=\sum\left(-\frac{1}{2}\ln(2\pi\theta_2)-\frac{(x_k-\theta_1)^2}{2\theta_2}\right)$$

$$\ln p(x_k|\theta)=-\frac{1}{2}\ln 2\pi\theta_2-\frac{1}{2\theta_2}(x_k-\theta_1)^2$$

$$\nabla_\theta l=\nabla_\theta \ln p(x_k|\theta)=\begin{bmatrix}\frac{1}{\theta_2}(x_k-\theta_1)\\ -\frac{1}{2\theta_2}+\frac{(x_k-\theta_1)^2}{2\theta_2^2}\end{bmatrix}$$

$$\frac{\partial l(\theta)}{\partial\theta_1}=\sum+\frac{1}{2\theta_2}2(x_k-\theta_1)=\sum_{k=1}^{n}\frac{x_k-\theta_1}{\theta_2}=0$$

$$\sum_{k=1}^{n}\frac{1}{\hat{\theta}_2}(x_k-\hat{\theta}_1)=0,\quad -\sum_{k=1}^{n}\frac{1}{\hat{\theta}_2}+\sum_{k=1}^{n}\frac{(x_k-\hat{\theta}_1)^2}{\hat{\theta}_2^2}=0,$$

$$\frac{\partial l(\theta)}{\partial\theta_2}=\sum-\frac{1}{2}\frac{1}{\cancel{2\pi}\theta_2}\cdot\cancel{2\pi}+\frac{(x_k-\theta_1)^2}{2}\left(\frac{1}{\theta_2^2}\right)$$

$$=\sum-\frac{1}{2\theta_2}+\frac{(x_k-\theta_1)^2}{2\theta_2^2}=0$$

Thus, $\hat{\theta}_1=\frac{1}{n}\sum_{k=1}^{n}x_k=\mu$, and $\hat{\theta}_2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{\mu})^2=\sigma^2$.

$$(\theta_2^{-1})'=-\theta_2^{-2}$$

$$\hat{\theta}_1=\frac{1}{n}\sum_{k=1}^{n}x_k$$

$$\hat{\theta}_2=\frac{1}{n}\sum_{k=1}^{n}(x_k-\hat{\theta}_1)$$