

The total point scatter is given by:

$$\begin{aligned}
 T &= \frac{1}{2} \sum_{i=1}^N \sum_{i'=1}^N d_{ii'} \quad , \text{ where } d_{ii'} = d(x_i, x_{i'}) \\
 &= \frac{1}{2} \sum_{h=1}^K \sum_{i: C(i)=h} \left(\sum_{i': C(i')=h} d_{ii'} + \sum_{i': C(i') \neq h} d_{ii'} \right) \\
 &= \underbrace{\frac{1}{2} \sum_{h=1}^K \sum_{i: C(i)=h} \sum_{i': C(i')=h} d_{ii'}}_{W(C)} + \underbrace{\frac{1}{2} \sum_{h=1}^K \sum_{i: C(i)=h} \sum_{i': C(i') \neq h} d_{ii'}}_{B(C)}
 \end{aligned}$$

$$\text{Let } d(x_i, x_{i'}) = d_{i,i'} = \|x_i - x_{i'}\|^2 \\ = (x_i - x_{i'})^T (x_i - x_{i'})$$

We have :

$$\begin{aligned} W(C) &= \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} d_{i,i'} \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (x_i - x_{i'})^T (x_i - x_{i'}) \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (x_i - \bar{x}_k + \bar{x}_k - x_{i'})^T (x_i - \bar{x}_k + \bar{x}_k - x_{i'}) \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (x_i - \bar{x}_k)^T (x_i - \bar{x}_k) \\ &\quad + \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (\bar{x}_k - x_{i'})^T (\bar{x}_k - x_{i'}) \\ &\quad \underbrace{\hspace{10em}}_{= (x_{i'} - \bar{x}_k)^T (x_{i'} - \bar{x}_k)} \\ &\quad + \frac{1}{2} \cdot 2 \sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (x_i - \bar{x}_k)^T (\bar{x}_k - x_{i'}) \end{aligned}$$

We have :

$$\begin{aligned} &\sum_{k=1}^K \sum_{i: C(i)=k} \sum_{i': C(i')=k} (x_i - \bar{x}_k)^T (\bar{x}_k - x_{i'}) \\ &= \sum_{k=1}^K \sum_{i: C(i)=k} (x_i - \bar{x}_k)^T \underbrace{\sum_{i': C(i')=k} (\bar{x}_k - x_{i'})}_{=0} \end{aligned}$$

$$\bar{x}_k = \frac{1}{N_k} \sum_{i: C(i)=k} x_i \\ N_k = \sum_{i=1}^N \mathbb{I}(C_i = k)$$

We get :

$$\begin{aligned} W(C) &= \frac{1}{2} \sum_{k=1}^K \sum_{i: C(i)=k} \overbrace{(x_i - \bar{x}_k)^T (x_i - \bar{x}_k)}^{\|x_i - \bar{x}_k\|^2} \underbrace{\sum_{i': C(i')=k} 1}_{= N_k} \\ &\quad + \frac{1}{2} \sum_{k=1}^K \sum_{i': C(i')=k} \underbrace{(x_{i'} - \bar{x}_k)^T (x_{i'} - \bar{x}_k)}_{= \|x_{i'} - \bar{x}_k\|^2} \underbrace{\sum_{i: C(i)=k} 1}_{= N_k} \\ &= \frac{1}{2} \sum_{k=1}^K N_k \sum_{i: C(i)=k} \|x_i - \bar{x}_k\|^2 + \frac{1}{2} \sum_{k=1}^K N_k \sum_{i': C(i')=k} \|x_{i'} - \bar{x}_k\|^2 \\ &= \sum_{k=1}^K N_k \sum_{i: C(i)=k} \|x_i - \bar{x}_k\|^2 \end{aligned}$$

$$\begin{aligned}
 \text{Let } g(m) &= \sum_{i \in S} \|x_i - m\|^2 \\
 &= \sum_{i \in S} (x_i - m)^T (x_i - m) \\
 &= \sum_{i \in S} x_i^T x_i + \underbrace{\sum_{i \in S} m^T m}_{m^T I_n} - 2 \sum_{i \in S} x_i^T m
 \end{aligned}$$

$\left. \begin{aligned} \frac{\partial a^T b}{\partial b} &= a \\ \frac{\partial b^T A b}{\partial b} &= (A + A^T) b \end{aligned} \right\}$

We have:

$$\frac{\partial g(m)}{\partial m} = \sum_{i \in S} \underbrace{(I + I^T)}_{2I} m - 2 \sum_{i \in S} x_i$$

$$= 2 \sum_{i \in S} m - 2 \sum_{i \in S} x_i$$

$$= 2 N_S m - 2 \sum_{i \in S} x_i, \quad N_S = \sum_{i=1}^N \mathbb{I}(i \in S)$$

To find the m that minimizes $g(m)$ we set

$$\frac{\partial g(m)}{\partial m} = 0 \quad \text{and solve for } m:$$

$$2 N_S m - 2 \sum_{i \in S} x_i = 0$$

$$\rightarrow m = \frac{1}{N_S} \sum_{i \in S} x_i = \bar{x}_S$$

We have:

$$\hat{C}_i = \operatorname{argmax}_h \left(\alpha_h p(x_i | C_i = h) \right)$$

$$= \operatorname{argmax}_h \left(\log(\alpha_h p(x_i | C_i = h)) \right)$$

$$= \operatorname{argmax}_h \left(\log(\alpha_h) + \log \left(\frac{1}{(2\pi)^{p/2} \sigma^p} e^{-\frac{1}{2\sigma^2} \|x_i - \mu_h\|^2} \right) \right)$$

$$= \operatorname{argmax}_h \left(\log(\alpha_h) - \frac{p}{2} \log(2\pi) - p \log(\sigma) - \frac{1}{2\sigma^2} \|x_i - \mu_h\|^2 \right)$$

$$= \operatorname{argmax}_h \left(\log(\alpha_h) - \frac{1}{2\sigma^2} \|x_i - \mu_h\|^2 \right)$$