Honor Code.

(a). Zheng Wu.

(b). I hereby state that all of my ~~following~~ solutions were entirely in my words and were written by me. I have not looked at another students solutions and I have fairly credited all external sources in this write up.

1.(a). $M = UDV^T$.

let matrix U_i be a matrix whose i-th column is the ith colum vector of matrix U and all the other colunms are all zero vectors.

$$U_i = \begin{bmatrix} 0 & \cdots & | & \cdots & 0 \\ 0 & & u_i & & 0 \\ \vdots & & | & & 0 \\ 0 & & & & 0 \end{bmatrix}$$

$\Rightarrow$ then $U = \sum_{i=1}^{n} U_i$ likewise $V = \sum_{i=1}^{n} V_i$

Since D is a diagonal matrix.

$\Rightarrow UD = \sum_{i=1}^{n} U_i \cdot d_i$

For any two matrixes A_i and B_k.

$A_i \cdot B_k^T = 0$ if $i \neq k$.

thus $UDV^T = (\sum_{i=1}^{n} U_i \cdot d_i)(\sum_{j=1}^{n} V_j^T)$

$= \sum_{i=1}^{n} U_i \cdot d_i V_j^T$

$= \sum_{i=1}^{n} U_i d_i V_i^T = \sum_{i=1}^{n} d_i U_i V_i^T$

1.(b).

$M = UDV^T$

① $M^TM = (UDV^T)^T(UDV^T)$

$= VD^TU^T \cdot U \cdot D \cdot V^T$

$= VD^TDV^T$

$\Rightarrow M^TM \cdot V = V \cdot D^T \cdot D \cdot V^T \cdot V = V \cdot D^T \cdot D$

thus the i-th eigenvalue of M^TM is d_i^2 and corresponding eigenvector is V_i.

② $MM^T = UDV^T(UDV^T)^T$

$= UDV^TVD^TU^T$

$= UDD^TU^T$

$MM^TU = UDD^TU^TU = UDD^T$

thus the i-th eigenvalue of MM^T is d_i^2 and corresponding eigenvector is U_i.

4.(a)

For the LHS:

$$\frac{1}{|C_k|}\sum_{i,i' \in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2 = \sum_i\sum_j x_{ij}^2 - 2\sum_i\sum_j x_{ij}\bar{x}_{kj} + \sum_i\sum_j x_{i'j}^2$$

$$= 2\sum_i\sum_j x_{ij}^2 - 2|C_k|\sum_j x_{kj}^2$$

For RHS:

$$2\sum_{i \in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 = 2\sum_i\sum_j x_{ij}^2 - 4\sum_i\sum_j x_{ij}\bar{x}_{kj} + 2\sum_i\sum_j \bar{x}_{kj}^2$$

$$= 2\sum_i\sum_j x_{ij}^2 - 2\left(2\sum_i\sum_j x_{ij}\bar{x}_{kj} - \sum_i\sum_j \bar{x}_{kj}^2\right)$$

$$= 2\sum_i\sum_j x_{ij}^2 - 2|C_k|\sum_j \bar{x}_{kj}^2$$

thus LHS = RHS

i.e. $$\frac{1}{|C_k|}\sum_{i,i' \in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2 = 2\sum_{i \in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2$$

4(b)

We know that $\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2 = 2\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2$ holds,

~~So~~ the cluster means for each feature are the constants that minimize the sum of squared deviations, ~~re~~ allocating the observations can only improve, thus the algorithm runs, ~~the~~

$$\sum_{k=1}^{K}\frac{1}{|C_k|}\sum_{i,i'\in C_k}\sum_{j=1}^{p}(x_{ij}-x_{i'j})^2$$ will always decrease. When results no longer changes, a local optimum has been reached.