Solution for the Exercise on Page 27 (Thursday Session)

Lemma 1 (Hessian determines Clustering). *Let R be the k-means risk function. Then the following holds for clusterings $\mathcal{C}, \mathcal{D}$. If the Hesse matrices of the risk functions $R_{\mathcal{C}}(w)$ and $R_{\mathcal{D}}(w)$, respectively, coincide at μ, then $\mathcal{C} = \mathcal{D}$.*

Proof. For sake of brevity, let

$$A_{p,q} := \left.\frac{\partial^2 R_{\mathcal{C}}(w)}{\partial w_p \partial w_q}\right|_{w=\mu} = (C_p = C_q) \cdot \frac{-2\langle x_p - z_C(w), x_q - z_C(w)\rangle}{S_C(w)} \ .$$

It suffices to show that centers $z_1, z_2, \ldots, z_k$ of partition $\mathcal{C}$ are uniquely determined by matrix A. To this end, we view A as the adjacency matrix of a graph G with nodes $x_1, x_2, \ldots, x_n$, where nodes x_p, x_q are connected by an edge if and only if $A_{p,q} \neq 0$. Let $K_1, K_2, \ldots, K_\ell$ be the connected components of G. Note that there is an edge between x_p and x_q only if p and q belong to the same cluster in $\mathcal{C}$. Thus, the connected components of G represent a refinement of partition $\mathcal{C}$. Consider a fixed cluster C_j in $\mathcal{C}$ with center z_j. Recall that

$$z_j = \sum_{x_i \in C_j} \mu_i x_i \ . \tag{1}$$

Let $K \subseteq C_j$ be any connected component of G that is contained in C_j and define, for sake of brevity, $\mu(K) := \sum_{x_i \in C} \mu_i$ and $K' = C_j \setminus K$. We claim that

$$z_j = \frac{1}{\mu(K)} \sum_{x_i \in K} \mu_i x_i \ , \tag{2}$$

that is, z_j is determined by any component $K \subseteq C_j$. Since this is obvious for $K = C_j$, we assume that $K \subset C_j$. We can rewrite (1) as

$$0 = \left(\sum_{x_i \in K} \mu_i (x_i - z_j)\right) + \left(\sum_{x_{i'} \in K'} \mu_{i'} (x_{i'} - z_j)\right) \ . \tag{3}$$

Pick any pair i, i' such that $x_i \in K$ and $x_{i'} \in K'$. Since x_i and $x_{i'}$ are not neighbors in G, $A_{i,i'} = 0$, which means that $x_i - z_j$ is orthogonal to $x_{i'} - z_j$. Thus the vector represented by the first sum in (3) is orthogonal on the vector represented by the second sum. It follows that both sums yield zero, respectively. Rewriting this for the first sum, we obtain (2). □