HWK 4

1. a) Let $f, g : D \to \mathbb{R}$ for some $D \subseteq \mathbb{R}^d$ and let $\mathbf{x}_0$ be an interior point of D. Assume f and g are continuously differentiable at $\mathbf{x}_0$. Let $h(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ for all $\mathbf{x}$ in D. Compute $\nabla h(\mathbf{x}_0)$. [*Hint:* You can make use of the corresponding single variable result.]

b) Let $f : D \to \mathbb{R}$ for some $D \subseteq \mathbb{R}^d$ and let $\mathbf{x}_0$ be an interior point of D where $f(\mathbf{x}_0) \neq 0$. Assume f is continuously differentiable at $\mathbf{x}_0$. Compute $\nabla(1/f)$. [*Hint:* You can make use of the corresponding single variable result.]

c) Let $f, g : D \to \mathbb{R}$ for some $D \subseteq \mathbb{R}^d$ and let $\mathbf{x}_0$ be an interior point of D. Assume f and g are twice continuously differentiable at $\mathbf{x}_0$. Let $h(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ for all $\mathbf{x}$ in D. Compute $\mathbf{H}_h(\mathbf{x}_0)$.

2. State *Taylor's Theorem* for the function

$$f(x_1, x_2) = \cos(x_1 x_2)$$

at $(x_1, x_2) = (0, 0)$. In particular, give explicit expressions for the gradient and Hessian terms. [*Hint:* You can use standard formulas for the derivatives of single-variable functions without proof.] $\lhd$

3. Fix a partition $C_1, \ldots, C_k$ of $[n]$. Under the k-means objective, its cost is

$$\mathcal{G}(C_1, \ldots, C_k) = \min_{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k \in \mathbb{R}^d} G(C_1, \ldots, C_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)$$

where

$$G(C_1, \ldots, C_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k) = \sum_{i=1}^{k} \sum_{j \in C_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 = \sum_{j=1}^{n} \|\mathbf{x}_j - \boldsymbol{\mu}_{c(j)}\|^2,$$

with $\boldsymbol{\mu}_i \in \mathbb{R}^d$, the center of cluster C_i.

a) Suppose $\boldsymbol{\mu}_i^*$ is a global minimizer of

$$F_i(\boldsymbol{\mu}_i) = \sum_{j \in C_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2,$$

for each i. Show that $\boldsymbol{\mu}_1^*, \ldots, \boldsymbol{\mu}_k^*$ is a global minimizer of $G(C_1, \ldots, C_k; \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k)$.

b) Compute the gradient of $F_i(\boldsymbol{\mu}_i)$.

c) Find the stationary points of $F_i(\boldsymbol{\mu}_i)$.

d) Compute the Hessian of $F_i(\boldsymbol{\mu}_i)$.

4. Consider the function

$$f(x_1, x_2) = (x_2 - x_1^2)(x_2 - 2x_1^2).$$

a) Compute the gradient and Hessian of f at $(0, 0)$.

b) Show that $(0, 0)$ is not a strict local minimizer of f. [*Hint:* Consider a parametric curve of the form $x_1 = t^\alpha$ and $x_2 = t^\beta$.]

c) Let $\mathbf{g}(t) = \mathbf{a}t$ for some nonzero vector $\mathbf{a} = (a_1, a_2)^T \in \mathbb{R}^2$. Show that $t = 0$ is a strict local minimizer of $h(t) = f(\mathbf{g}(t))$.