MATH 462 ASSIGNMENT 1
VERSION September 27, 2022

ADAM M. OBERMAN

1.1. **Instructions.** Refer to class notes and typed notes posted on `https://adam-oberman.github.io/Math462/`, as well as textbook references.

Due: Tuesday Sept 27. Submit your solutions on MyCourses course page.

Math Solutions should be *handwritten*. (If you insist on Latex, get my permission). You can get help from other students, but you should do the write up yourself. Code solutions can be a PDF of a workbook, you can use any language.

1.2. **Vector Calculus for ML.**

Exercise 1.1. Let $f : \mathbb{R}^d \to \mathbb{R}$ and write $\|x\| = \sqrt{x_1^2 + \cdots + x_d^2}$. Find $\nabla f(x)$ for each of the following functions. Write the answer in vector notation.

(1) $f(x) = \|x\|^2$.
(2) $f(x) = \|x\|$,
(3) $f(x) = \|x - a\|$
(4) $f(x) = 1/\|x\|$
(5) $f(x) = \|m \cdot x - b\|^2$, where $m \in \mathbb{R}^d, b \in \mathbb{R}$
(6) $f(x) = \|m_1 \cdot x - b_1\|^2 + \dots \|m_n \cdot x - b_n\|^2$, where $m_i \in \mathbb{R}^d, b_i \in \mathbb{R}$

Exercise 1.2. (This question was done in class notes in the case $d = 1$). Let $w, x_1, \dots, x_m \in \mathbb{R}^d$. Set

$$\widehat{L}(w) = f(w, x_1, \dots, x_m) = \frac{1}{m} \sum_{i=1}^{m} \|w - x_i\|_2^2$$

Show

$$\nabla_w f(w, x_1, \dots, x_m) = 2w - 2\bar{x}, \quad \bar{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$$

and

$$w^* = \arg\min_w \widehat{L}(w) = \bar{x}, \qquad \widehat{L}(w^*) = \frac{1}{m} \sum_{i=1}^{m} \|\bar{x} - x_i\|^2$$

Exercise 1.3 (Derive Matrix equation).
(1) Find $\nabla_x f(x)$ where $f(x) = \|Mx - b\|^2$, where M is $n \times d$ and $b \in \mathbb{R}^n$. Write you answer in matrix notation.
(2) Find $\nabla_w \widehat{L}(w)$, where $\widehat{L}(w) = \frac{1}{m}\|Xw - Y\|^2$ and X is $n \times d$, and $Y \in \mathbb{R}^n$. Write your answer in matrix notation. Hint: this is just a notation change from the previous problem.
(3) Show that a critical point of $\widehat{L}$ is characterized by the solution of the linear system $X^T X w = X^T Y$

Note: This result derived (using different notation) in MML Example 5.11.

1.3. **k-means Clustering.**

Exercise 1.4. For the following two dimensional distributions, sketch by hand a few points sampled from each distribution. using the instructions as a guideline (this should be a simple sketch). Indicate on the sketch the average distance in each cluster.

(1) Consider 4 normal distributions, with means μ_i on the vertices $(\pm 1, \pm 1)$ of a square of side length 2, each with $\sigma_i = \sigma = .01$. Indicate the 4-means clustering.
(2) Consider the uniform distribution on the 4 by 2 rectangle with vertices $(\pm 2, \pm 1)$. Indicate using two plots two different ways to cluster with 2-means.

Exercise 1.5. **This one has been moved to HW, so we can go over the programming background** ~~Now we will repeat the last exercise using a code implementation. In each case, generate a few dozens points from each distribution (enough to make the plot clear). Then run the k-means algorithm on the data. Plot the distribution of points along with the result of the algorithm by indicating the means μ_i of the clusters. Also write or indicate graphically, the σ_i (which is the square root of the mean of the distances squared) for each cluster~~

(1) Consider 4 normal distributions, with means μ_i on the vertices $(\pm 1, \pm 1)$ of a square of side length 2, each with $\sigma_i = \sigma = 1$. Indicate the 4-means clustering.

(2) Consider the uniform distribution on the 4 by 2 rectangle with vertices $(\pm 2, \pm 1)$. Indicate using two plots two different ways to cluster with 2-means.

Exercise 1.6. Plot $h_W(x)$ in dimension 1, when $W = (2, 5, 9)$. How does the function plot change for $W = (2, 8, 9)$?

Exercise 1.7. Consider the k-means algorithm as defined in the posted notes.

(1) For each fixed j, the algorithm defines $w_j^{t+1} = \arg\min_{w \in \mathbb{R}^d} \sum_{x \in C_j} \|x - w\|^2$. Prove that w_j^{t+1} is the mean of the cluster members, in other words,

$$w_j^{t+1} = \frac{1}{|C_j|} \sum_{x \in C_j} x$$

In addition, show that

$$\sum_{x \in C_j} \|x - w_j^{t+1}\|^2 \leq \sum_{x \in C_j} \|x - w_j^t\|^2$$

with strict equality, unless $w_j^{t+1} = w_j^t$.

(2) (Proof of Lemma 3.5 from the posted notes) Suppose we update h_W according to the k-means algorithm (see notes), and let W^t be the corresponding weights. Prove that,

$$\widehat{L}(h_W^{t+1}) \leq \widehat{L}(h_W^t)$$

with a strict inequality, unless $W^{t+1} = W^t$.