Homework 3
MATH 123 - Spring 2023
Tufts University, Department of Mathematics
Due: February 9, 2023

Question 1

Let $x, y \in \mathbb{R}^{d\times 1}$. Prove that $xy^T \in \mathbb{R}^{d\times d}$ is rank 1.

Question 2

Prove that the Euclidean dot product $\langle x, y\rangle = \sum_{i=1}^n x_i y_i,\ x, y \in \mathbb{R}^n$ is an inner product, where an *inner product* is a function $\langle\cdot,\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that:

(a) For all $x, y \in \mathbb{R}^n$, $\langle x, y\rangle = \langle y, x\rangle$.
(b) For all $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$.
(c) For all $x, y, z \in \mathbb{R}^n$, $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
(d) For all $x \in \mathbb{R}^n$, $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only if $x = 0$.

Question 3

Suppose M is a symmetric $d \times d$ matrix and $x, y \in \mathbb{R}^{d\times 1}$.

(a) Prove that $\langle x, y\rangle_M = x^\top M y$ is an inner product if M is positive-definite.
(b) Prove that $\langle x, y\rangle_M$ as above need not be an inner product if M is only positive semi-definite.

Question 4

Let $x_1, \ldots x_n \subset \mathbb{R}^d$. Fix some positive integer K. Let $C_1, \ldots, C_K$ be a partition of the data with centroids $\mu_1, \ldots, \mu_K$. Let

$$F(C_1, \ldots, C_K) = \sum_{k=1}^K \sum_{x_i \in C_k} \|\mu_k - x_i\|_2^2$$

(a) Prove that, for a fixed K, F achieves a minimum value.
(b) What is the minimum value if $K = n$?

Question 5

Run the MATLAB script 'Kmeans_Gaussians'.

(a) Run K-means with $K = 2$, 100 replicates. Show the output visually.
(b) Plot the error of the K-means functional as a function of the number of iterations. Is there convergence?
(c) Do the clusters accord with your intuition?

1. Question 6

Run the MATLAB script 'Kmeans_Ellipses'.

(a) Run K-means with $K = 2$, 100 replicates. Show the output visually.
(b) Plot the error of the K-means functional as a function of the number of iterations. Is there convergence?
(c) Do the clusters accord with your intuition?