Homework 3
MATH 123 - Fall 2018
Tufts University, Department of Mathematics
Due: September 25, 2018

1. Question 1

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

2. Question 2

Let $A \in \mathbb{R}^{m\times n}$ have largest singular value σ_1. Prove that

$$\sigma_1 = \max_{x\in\mathbb{R}^n, \|x\|_2=1} \|Ax\|_2.$$

3. Question 3

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 $x = 0$.

4. Question 4

(a) Prove that $\langle x, y\rangle_M = xMy^T$ 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.

5. Question 5

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

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

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

6. Question 6

Run the MATLAB script 'Kmeans_Gaussians'.

(a) Run k-means with 100 replicates.
(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?

7. Question 7

Run the MATLAB script 'Kmeans_Ellipses'.

(a) Run k-means with 100 replicates.
(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?