

Homework #3

Eric Tao
Math 123: Homework #3

February 3, 2023

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

Solution. First, assume we're not in the degenerate case $x = 0$ or $y = 0$, as if either are true, then $xy^T = 0$, the 0 matrix.

Here, we recall that, by definition:

$$(xy^T)_{ij} = \sum_{k=1}^1 x_{ik}y_{kj} = x_iy_j$$

where I drop the second index because, of course, these are vectors. Fix some i . Then, looking at the i -th row, we notice that we may factor x_i from every term. In particular, as a vector, the i -th row looks like:

$$(x_iy_1 \quad x_iy_2 \quad \dots \quad x_iy_d) = x_i(y_1 \quad y_2 \quad \dots \quad y_d)$$

where we notice that the i -th row is a multiple of y , as a row vector. Since the choice of i was completely arbitrary, this process can be done for every row, and thus every row vector is a multiple of y . Thus, the row space has dimension 1, and since we're a square matrix, we have that the dimension of the row space is equal to that of the column space, and the rank is 1.

□

Question 2. Prove that the Euclidean dot product $\langle x, y \rangle = \sum_{i=1}^n x_iy_i$, $x, y \in \mathbb{R}^n$ is an inner product, where an inner product is a binary function from a (real-valued) vector space V to a field F , $\langle \cdot, \cdot \rangle : V \times V \rightarrow F$ such that the following hold (in the context of a real vector space):

- (a) For all $x, y \in V$, $\langle x, y \rangle = \langle y, x \rangle$
- (b) For all $x, y \in V$, $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
- (c) For all $x, y, z \in V$, $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
- (d) For all $x \in V$, $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \iff x = 0$

Solution. (a)

This should be clear by the commutativity of real numbers:

$$\langle x, y \rangle = \sum_{i=1}^n x_iy_i = \sum_{i=1}^n y_ix_i = \langle y, x \rangle$$

(b)

This should be clear by how multiplication distributes over addition:

$$\langle \alpha x, y \rangle = \sum_{i=1}^n (\alpha x_i)y_i = \alpha \sum_{i=1}^n y_ix_i = \alpha \langle y, x \rangle$$

(c)

Same as (b), follows from distributive property of multiplication:

$$\langle x + y, z \rangle = \sum_{i=1}^n (x_i + y_i) z_i = \sum_{i=1}^n x_i z_i + \sum_{i=1}^n y_i z_i = \sum_{i=1}^n x_i z_i + \sum_{j=1}^n y_j z_j = \langle x, z \rangle + \langle y, z \rangle$$

(d)

First, we consider the expansion of $\langle x, x \rangle$:

$$\langle x, x \rangle = \sum_{i=1}^n x_i x_i = \sum_{i=1}^n x_i^2$$

Since we have that for all $x_i \in \mathbb{R}$, $x_i^2 \geq 0$, we have that $\sum_{i=1}^n x_i^2$ is a sum of non-negative numbers, and thus must be at least 0. Thus, $\langle x, x \rangle \geq 0$

It is obvious that if $x = 0$, then $\langle x, x \rangle = \sum_{i=1}^n 0 * 0 = 0$. Now, suppose $\langle x, x \rangle = 0$. Then, we have that $\sum_{i=1}^n x_i^2 = 0$. Since, again, these are non-negative numbers, this can only be 0 if $x_i = 0$ for all i . But, if $x_i = 0$ for all i , x is the 0 vector. □

Question 3. (a) Prove that $\langle x, y \rangle_M = xMy^T$ satisfies the properties of an inner product if M is positive definite.

(b) Show that $\langle x, y \rangle_M$ need not be an inner product if M is positive semi-definite.

Solution. (a)

Suppose M is positive definite. We check each property in turn:

$$\langle x, y \rangle_M = xMy^T = (xMy^T)^T = yM^T x^T = yMx^T = \langle y, x \rangle_M$$

where we use the fact that since xMy^T is a scalar, and thus may be interpreted as a 1×1 matrix, it must be symmetric, and that M being positive-definite means that M is symmetric.

$$\langle \alpha x, y \rangle_M = (\alpha x)My^T = \alpha(xMy^T) = \alpha \langle x, y \rangle_M$$

where we use the fact that we can pull scalars out from matrix multiplication.

$$\langle x + y, z \rangle_M = (x + y)Mz^T = xMz^T + yMz^T = \langle x, z \rangle_M + \langle y, z \rangle_M$$

where we use the fact that matrix multiplication distributes over matrix addition.

and the last property comes by the definition of positive definite, as by definition, $xMx^T \geq 0$ for all x , and is 0 if and only if $x = 0$. So we notice that:

$$\langle x, x \rangle_M = xMx^T \geq 0$$

and

$$\langle x, x \rangle_M = 0 \iff xMx^T = 0 \iff x = 0$$

So, $\langle x, y \rangle_M$ defines an inner product.

(b)

We show a counter example. Consider the matrix:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

It should be clear that this matrix is positive semi-definite because we may compute the characteristic polynomial and eigenvalues as $(1 - \lambda)\lambda = 0 \implies \lambda = 0, 1 \geq 0$, and since A is symmetric, with non-negative eigenvalues, it must be positive semi-definite.

However, this matrix does not define an inner product. Consider the vector $v = \begin{pmatrix} 0 & 1 \end{pmatrix}$. We have that:

$$vAv^T = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0$$

Thus, because we found a non-0 v such that $vAv^T = 0$, we have that $vAv^T = 0 \not\Rightarrow v = 0$, and thus we do not satisfy the biconditional. □

Question 4. Let $x_1, \dots, x_n \in \mathbb{R}^d$. Fix some positive integer K . Let C_1, \dots, C_K be a partition of the data with centroids μ_1, \dots, μ_K . Let

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

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

Solution. (a)

Fix some $K \in \mathbb{N}$. Since n is a fixed natural number, we have finitely many data points. In particular, if we have finitely many data points, then we can only have finitely many partitions of n into K subsets. Then, if we consider the values of F_i , where we index a partition Γ_i by i , this is a set of non-negative real numbers, finite. By the least upper bound property of the real numbers, this must have an infimum since this is bounded below by 0, and further, since this is a finite set, it must achieve the infimum, since finite sets do not have limit points. Thus, F must achieve a minimum for fixed values of K .

(b)

If $K = n$, then trivially, the only partition allowable is $C_1 = \{x_1\}, C_2 = \{x_2\} \dots C_n = \{x_n\}$ up to relabeling, since we have n data points and n boxes. Then, the centroid of each is exactly $\mu_j = x_j$, so then we have that for each C_n , that

$$\sum_{x_i \in C_j} \|\mu_j - x_i\|_2^2 = \|x_j - x_j\|_2^2 = 0$$

Thus, $F = \sum_{K=1}^n 0 = 0$, and the minimum and only value of F if $K = n$ is 0. □

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 agree with your intuition?

Solution. □

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 agree with your intuition?

Solution.

