

The Inner Products AA^T and $A^T A$

Ng Yen Kaow

Notations

- A vector is by default a column
 - For vectors x and y , inner (or dot) product
$$\langle x, y \rangle = x^T y$$
 - $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle = x^T y + z^T y$
 - Some texts use row vectors and $\langle x, y \rangle = xy^T$
- For a matrix
 - Each row represents an example or datapoint while each column is a feature
 - Features are like fixed columns in a spreadsheet
 - For matrices X and Y , $\langle X, Y \rangle = XY^T$ or $(x_i y_i^T)$
 - Some texts define columns as examples and use $X^T Y$ as the inner product
- Just write $x^T x$, $x^T M x$, and XX^T , $Q\Lambda Q^T$, K?

What about outer products?

- The outer product of two vectors is a matrix

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix}$$

- The outer product (or Kronecker product) of two matrices is a **tensor**
- We don't deal with outer products yet

Python call for inner product

- Inner products are performed with `np. dot ()`
 - When called on two arrays, the arrays are **automatically** oriented to perform inner product
 - But note that `[[1], [1]]` is a 1×2 matrix
 - When called on an array x and a matrix X , the array is **automatically** read as a row for `np. dot (x, X)`, and column for `np. dot (X, x)` to perform inner product
 - When called on two matrices, make sure that the matrices are oriented correctly, or you will get $X^T X$ when you want XX^T
 - You can't get an outer product with `np. dot ()`
- If you write `x*y` or `X*Y`, what you get is an element-wise multiplication

The inner products $A^T A$ and AA^T

- Given an $n \times m$ matrix A where the rows are datapoints and columns are features
- The inner product $A^T A$ is the **covariance matrix**
 - Used in the proof of PCA
- The inner product AA^T is called a **Gram matrix**, or **Gramian**
 - Used in the proof of MDS
- Eigendecomposition of both $A^T A$ and AA^T are equivalent (convertible from each other)

Properties of AA^T (or $A^T A$)

□ Properties

- Positive semi-definite (proof next slide)
- Equivalent eigendecomposition (proof later)

□ AA^T and $A^T A$ are related

- Convertible through their eigenvectors (proof later)
- Obtainable from SVD of A

AA^T/A^TA is positive semi-definite

- A matrix M is positive semi-definite iff all its eigenvalues are non-negative

- That is, $\forall x (x^T M x \geq 0)$

- For example, $M = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ is positive semi-definite because

$$\begin{aligned} (x_1 \quad x_2) \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= x_1^2 - 2x_1x_2 + x_2^2 \\ &= (x_1 - x_2)^2 \geq 0 \end{aligned}$$

- To show Gramians are PSD, we first establish the equivalence between $x^T M x$ and a quadratic formula

Quadratic form

- A generalized quadratic formula of n variables can be written in the form of $x^T M x$
- For instance, a quadratic formula of two variables

$$a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2$$

can be written as

$$(x_1 \quad x_2) \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= (x_1 a_{11} + x_1 a_{12} \quad x_1 a_{12} + x_2 a_{22}) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2^2$$

- The general form of n variables is

$$(x_1 \quad \dots \quad x_n) \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x^T A x = \sum_{ij} a_{ij} x_i x_j$$

AA^T in quadratic form

- Let $B = AA^T$, then $x^T B x \geq 0$, let a_i be the i -row of A , then

$$\begin{aligned}x^T (AA^T) x &= \sum_{ij} \langle a_i, a_j \rangle x_i x_j \\&= \sum_{ij} \langle x_i a_i, x_j a_j \rangle \\&= \langle \sum_i x_i a_i, \sum_j x_j a_j \rangle \geq 0\end{aligned}$$

- This says that $x^T (AA^T) x$ can be factorized into a linear addition of the terms

$$(\sum_k x_i a_{ik})^2$$

- Hence AA^T is positive semi-definite (and similarly shown for $A^T A$)

Example of 2×2 matrix

$$A = \begin{pmatrix} \leftarrow a_1 \rightarrow \\ \leftarrow a_2 \rightarrow \end{pmatrix}, A^T = \begin{pmatrix} \uparrow & \uparrow \\ a_1^T & a_2^T \\ \downarrow & \downarrow \end{pmatrix}$$

$$AA^T = \begin{pmatrix} a_1 a_1^T & a_1 a_2^T \\ a_2 a_1^T & a_2 a_2^T \end{pmatrix}$$

$$x^T AA^T x = (x_1 \quad x_2) \begin{pmatrix} a_1 a_1^T & a_1 a_2^T \\ a_2 a_1^T & a_2 a_2^T \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= (x_1 a_1 a_1^T + x_2 a_2 a_1^T \quad x_1 a_1 a_2^T + x_2 a_2 a_2^T) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= a_1 a_1^T x_1^2 + 2a_1 a_2^T x_1 x_2 + a_2 a_2^T x_2^2 \quad (\because a_1 a_2^T = a_2 a_1^T)$$

$$= (x_1 a_1 + x_2 a_2)(x_1 a_1^T + x_2 a_2^T)$$

$$= \|x_1 a_1 + x_2 a_2\|^2$$

$$\geq 0$$

Example of 2×2 matrix

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

$$\begin{aligned} x^T A A^T x &= (x_1 \quad x_2) \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= (x_1 + x_2 \quad x_1 + 2x_2) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= x_1^2 + x_1 x_2 + x_1 x_2 + 2x_2^2 \\ &= (x_1^2 + 2x_1 x_2 + x_2^2) + x_2^2 \\ &= (x_1 + x_2)^2 + x_2^2 \end{aligned}$$

By theorem, $\sum_i x_i a_i = x_1(1 \ 0) + x_2(1 \ 1) = (x_1 + x_2 \quad x_2)$

$$\begin{aligned} (\sum_i x_i a_i)(\sum_i x_i a_i) &= (x_1 + x_2 \quad x_2) \begin{pmatrix} x_1 + x_2 \\ x_2 \end{pmatrix} \\ &= (x_1 + x_2)^2 + x_2^2 \end{aligned}$$

$$x^T A A^T x = (x_1 + x_2)^2 + x_2^2 \geq 0$$

Example of 3×2 matrix

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

$$\begin{aligned} x^T A A^T x &= (x_1 \quad x_2 \quad x_3) \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= (x_1 + x_3 \quad x_2 + x_3 \quad x_1 + x_2 + 2x_3) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \\ &= x_1^2 + 2x_1x_3 + x_2^2 + 2x_2x_3 + 2x_3^2 \end{aligned}$$

$$\begin{aligned} \text{By theorem, } \sum_i x_i a_i &= x_1(1 \quad 0) + x_2(0 \quad 1) + x_3(1 \quad 1) \\ &= (x_1 + x_3 \quad x_2 + x_3) \end{aligned}$$

$$\begin{aligned} (\sum_i x_i a_i)(\sum_i x_i a_i) &= (x_1 + x_3 \quad x_2 + x_3) \begin{pmatrix} x_1 + x_3 \\ x_2 + x_3 \end{pmatrix} \\ &= x_1^2 + 2x_3^2 + 2x_1x_3 + x_2^2 + 2x_2x_3 \end{aligned}$$

$$x^T A A^T x = (x_1 + x_3)^2 + (x_2 + x_3)^2 \geq 0$$

AA^T and $A^T A$ equal decomposition

- AA^T and $A^T A$ have equivalent eigendecomposition
- We will prove three facts
 1. AA^T and $A^T A$ have the same rank
 - 2.1 AA^T and $A^T A$ have the same eigenvalues
 - 2.2 AA^T and $A^T A$ have the same eigenvectors (different only to a transformation)

AA^T and $A^T A$ have the same rank

- Let $N(A)$ denote the null space of A
 - $N(A) = \{x | Ax = 0\}$
 - For $u \in N(A)$, $Au = 0 \Rightarrow A^T Au = 0 \Rightarrow u \in N(A^T A)$
 - For $u \in N(A^T A)$,
 $A^T Au = 0 \Rightarrow uA^T Au = 0$
 $\Rightarrow (Au)^T (Au) = 0 \xRightarrow{\text{See next slide}} Au = 0$
 $\Rightarrow u \in N(A)$
 - Hence $N(A^T A) = N(A) \Rightarrow \text{rank}(A^T A) = \text{rank}(A)$
 - Similarly $N(AA^T) = N(A^T)$
 $\Rightarrow \text{rank}(AA^T) = \text{rank}(A^T)$
- $\therefore \text{rank}(A) = \text{rank}(A^T), \text{rank}(A^T A) = \text{rank}(AA^T)$

Proof $X^T X = 0 \Rightarrow X = 0$

- Let $X = (x_{ij})$
- Observe that $(X^T X)_{ij} = \sum_k x_{ik} x_{jk}$
 $\Rightarrow (X^T X)_{ii} = \sum_k x_{ik}^2$
- Now $X^T X = 0 \Rightarrow \forall i, (X^T X)_{ii} = 0$
 $\Rightarrow \forall i \forall k, x_{ik} = 0$

AA^T and $A^T A$ equal eigenpairs

- **For any matrices A and B , AB and BA have the same non-zero eigenvalues**
 - Let $\lambda \neq 0$ be a eigenvalue for AB with eigenvector v
 - Then $ABv = \lambda v \Rightarrow BABv = \lambda Bv$
 $\Rightarrow (BA)(Bv) = \lambda(Bv)$
 $\Rightarrow \lambda$ is a eigenvalue of BA with eigenvector (Bv)
- **This result holds for all A and B**

Convert $AA^T \leftrightarrow A^T A$

- Let Λ be the diagonal matrix of eigenvalues for both AA^T and $A^T A$
- Let P, Q be their respective eigenvectors: $AA^T U = \Lambda U$ and $A^T A V = \Lambda V$, then

$$\begin{aligned} AA^T &= U \Lambda U^T \\ &= U (V^T A^T A V) U^T \\ &= (U V^T) A^T A (V U^T) \end{aligned}$$

AA^T and $A^T A$ related through SVD

- Singular values and vectors of AA^T and $A^T A$ are related to the singular values and vectors of A
- Let $A = UDV^T$ (UDV^T is the SVD of A), then

$$A^T A = VD^2V^T, \text{ and}$$

$$AA^T = UD^2U^T$$

(Proof omitted)

Kernel Gramian

- A kernel is a function that computes an inner product $d(x_1, x_2)$ without knowing X
 - X will be Euclidean but XX^T