Gramian and Covariance

Ng Yen Kaow

The inner products $AA^\top$ and $A^\top A$

- Given an $n \times m$ matrix A where the rows are datapoints, columns are features with zero sum for each column
  - The inner product $A^\top A$ is the covariance matrix (more precisely $A^\top A/n$)
    - Used in PCA
  - The inner product $AA^\top$ is called a Gram matrix, or Gramian
    - Used in Multidimensional Scaling
    - Used in Kernel method
  - Eigendecomposition of $A^\top A$ and $AA^\top$ are closely related

Properties of $AA^\top$ and $A^\top A$

- $AA^\top$ and $A^\top A$ are related
  - Closely related eigendecomposition (later)
  - Simultaneous solution in SVD (shown)

- $M = AA^\top \Leftrightarrow \forall x, x^\top M x \geq 0$ ($\Rightarrow$) proof later

  Similarly $M = A^\top A \Leftrightarrow \forall x, x^\top M x \geq 0$
  - M where $(\forall x)[x^\top M x \geq 0]$ holds is said to be positive semi-definite (PSD)
  - Kernel matrices (in kernel method) need to be positive semi-definite
  - A is called the square root of M

$M = AA^{\mathsf{T}} \Rightarrow M$ is PSD

- Given M, to show M is positive semi-definite, need to show $\forall x(x^{\mathsf{T}}Mx \geq 0)$
- For example, $M = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}$ is positive semi-definite because

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

- To show that AA^{T} is positive semi-definite, we first establish the equivalence between $x^{\mathsf{T}}Mx$ and a quadratic formula

$M = AA^\top \Rightarrow M$ is PSD

- A generalized quadratic formula of n variables can be written in the form of $x^\top M x$
- 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

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

$$= \begin{pmatrix} x_1a_{11} + x_1a_{12} & x_1a_{12} + x_2a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

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

- The general form of n variables is

$$\begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x^\top A x = \sum_{ij} a_{ij}x_ix_j$$

$M = AA^\top \Rightarrow M$ is PSD

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

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

- This says that $x^\top(AA^\top)x$ can be factorized into a linear addition of the terms t^2 for each element t in the vector $\sum_i a_i x_i$
  - Hence $AA^\top$ is positive semi-definite

PSD. Example of 2×2 matrix

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

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

$$\begin{aligned}
x^\top AA^\top x &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_1 a_1^\top & a_1 a_2^\top \\ a_2 a_1^\top & a_2 a_2^\top \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= \begin{pmatrix} x_1 a_1 a_1^\top + x_2 a_2 a_1^\top & x_1 a_1 a_2^\top + x_2 a_2 a_2^\top \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\
&= a_1 a_1^\top x_1^2 + 2 a_1 a_2^\top x_1 x_2 + a_2 a_2^\top x_2^2 \qquad (\because a_1 a_2^\top = a_2 a_1^\top) \\
&= (x_1 a_1 + x_2 a_2)(x_1 a_1^\top + x_2 a_2^\top) \\
&= \|x_1 a_1 + x_2 a_2\|^2 \\
&\geq 0
\end{aligned}$$

PSD. Example of 2×2 matrix

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

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

By theorem, $\sum_i a_i x_i = x_1 \begin{pmatrix} 1 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 1 & 1 \end{pmatrix} = \begin{pmatrix} x_1 + x_2 & x_2 \end{pmatrix}$

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

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

PSD. Example of 3×2 matrix

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

$$x^\top A A^\top x = \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= \begin{pmatrix} x_1 + x_3 & x_2 + x_3 & x_1 + x_2 + 2x_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= x_1^2 + 2x_1x_3 + x_2^2 + 2x_2x_3 + 2x_3^2$$

By theorem, $\sum_i a_i x_i = x_1 \begin{pmatrix} 1 & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 1 \end{pmatrix} + x_3 \begin{pmatrix} 1 & 1 \end{pmatrix}$

$$= \begin{pmatrix} x_1 + x_3 & x_2 + x_3 \end{pmatrix}$$

$$\left(\sum_i a_i x_i\right)\left(\sum_i a_i x_i\right) = \begin{pmatrix} x_1 + x_3 & x_2 + x_3 \end{pmatrix} \begin{pmatrix} x_1 + x_3 \\ x_2 + x_3 \end{pmatrix}$$

$$= x_1^2 + 2x_3^2 + 2x_1x_3 + x_2^2 + 2x_2x_3$$

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

$AA^\top / A^\top A$ equal in decomposition

- $AA^\top$ and $A^\top A$ have equivalent eigen-decomposition
- We will prove these facts

1. $AA^\top$ and $A^\top A$ have the same rank

2.1 $AA^\top$ and $A^\top A$ have the same eigen-values

2.2 $AA^\top$ and $A^\top A$ eigenvectors related (different for only up to an orthogonal transformation)

$AA^\top$ and $A^\top A$ have the same rank

- Let $N(A)$ denote the null space of A
  - $N(A) = \{x | Ax = 0\}$
- For $u \in N(A)$, $Au = 0 \Rightarrow A^\top Au = 0 \Rightarrow u \in N(A^\top A)$
- For $u \in N(A^\top A)$,

  $A^\top Au = 0 \Rightarrow uA^\top Au = 0$

  $\Rightarrow (Au)^\top (Au) = 0 \xRightarrow{\text{See next slide}} Au = 0$

  $\Rightarrow u \in N(A)$
- Hence $N(A^\top A) = N(A) \Rightarrow \text{rank}(A^\top A) = \text{rank}(A)$
- Similarly $N(AA^\top) = N(A^\top)$

  $\Rightarrow \text{rank}(AA^\top) = \text{rank}(A^\top)$

$\because \text{rank}(A) = \text{rank}(A^\top)$, $\text{rank}(A^\top A) = \text{rank}(AA^\top)$

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

- Let $X = (x_{ij})$
- First note that $X^\top X = 0$ is a system of equations where each element of $X^\top X$

$$(X^\top X)_{ij} = \sum_k x_{ki} x_{kj} = 0$$

- Since $(X^\top X)_{ii} = \sum_k x_{ki}^2 = 0$

$$\Rightarrow \forall i \forall k, x_{ki} = 0$$

$AA^\mathsf{T}/A^\mathsf{T}A$ have equal eigenvalues

(Stronger proof than that shown with SVD)

- **For any matrices A and B**, AB and BA have the same non-zero eigen-values
  - 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 eigen-vector (Bv)

$AA^\top / A^\top A$ eigenvectors related

- *(From previous slide)*

  Let $\lambda \neq 0$ be a eigenvalue for AB with eigenvector v

  $\Rightarrow \lambda$ is a eigenvalue of BA with eigen-vector (Bv)
- Let $A = X^\top$ and $B = X$, if v is eigen-vector for $X^\top X$, Xv is eigenvector for $XX^\top$
  - This is the basis for the equivalence of PCA and cMDS (later)

Convert $AA^\top \leftrightarrow A^\top A$

- Let Λ be the diagonal matrix of eigen-values for both $AA^\top$ and $A^\top A$
- Let U, V be their respective eigen-vectors, that is, $AA^\top U = \Lambda U$ and $A^\top AV = \Lambda V$, then

$$
\begin{aligned}
AA^\top &= U\Lambda U^\top \\
&= U(V^\top A^\top AV)U^\top \\
&= (UV^\top)A^\top A(VU^\top)
\end{aligned}
$$