AA^{T} and $A^{\mathrm{T}}A$ (Gramian and Covariance)

Ng Yen Kaow

The inner products AA^{T} and $A^{\mathrm{T}}A$

- Given an $n \times m$ matrix A where the rows are datapoints and columns are features
  - The inner product $A^{\mathrm{T}}A$ is the covariance matrix (more precisely $A^{\mathrm{T}}A/n$)
    - 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 $A^{\mathrm{T}}A$ and AA^{T} are equivalent (convertible from each other)

Properties of AA^{T} and $A^{\mathrm{T}}A$

- Properties
  - Positive semi-definite (proof later)
    - Furthermore, positive semi-definiteness of symmetric matrices is preserved over sum, product, and scaling
- AA^{T} and $A^{\mathrm{T}}A$ are related
  - Equivalent eigendecomposition (later)
  - Convertible through their eigenvectors (later)
  - Obtainable from SVD of A (proof omitted)

$AA^{\mathrm{T}}/A^{\mathrm{T}}A$ is positive semi-definite

- A matrix M is positive semi-definite (PSD) iff all its eigenvalues are non-negative
  - That is, $\forall x\left(x^{\mathrm{T}}Mx \geq 0\right)$
- 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 PSD, we first establish the equivalence between $x^{\mathrm{T}}Mx$ and a quadratic formula

Quadratic form

- A generalized quadratic formula of n variables can be written in the form of $x^{\mathrm{T}}Mx$
- 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
$$\begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x^{\mathrm{T}}Ax = \sum_{ij} a_{ij}x_ix_j$$

AA^{T} in quadratic form

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

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

- This says that $x^{\mathrm{T}}(AA^{\mathrm{T}})x$ can be factorized into a linear addition of the terms $\left(\sum_i x_i a_{ik}\right)^2$
  - Hence AA^{T} is PSD (and similarly so is $A^{\mathrm{T}}A$)

Example of 2×2 matrix

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

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

$$\begin{aligned} x^{\mathrm{T}} A A^T x &= \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_1 a_1^{\mathrm{T}} & a_1 a_2^{\mathrm{T}} \\ a_2 a_1^{\mathrm{T}} & a_2 a_2^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} x_1 a_1 a_1^{\mathrm{T}} + x_2 a_2 a_1^{\mathrm{T}} & x_1 a_1 a_2^{\mathrm{T}} + x_2 a_2 a_2^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= a_1 a_1^{\mathrm{T}} x_1^2 + 2 a_1 a_2^{\mathrm{T}} x_1 x_2 + a_2 a_2^{\mathrm{T}} x_2^2 \qquad (\because a_1 a_2^{\mathrm{T}} = a_2 a_1^{\mathrm{T}}) \\ &= (x_1 a_1 + x_2 a_2)\left(x_1 a_1^{\mathrm{T}} + x_2 a_2^{\mathrm{T}}\right) \\ &= \|x_1 a_1 + x_2 a_2\|^2 \\ &\geq 0 \end{aligned}$$

Example of 2×2 matrix

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

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

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

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

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

Example of 3×2 matrix

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

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

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

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

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

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

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

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

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

$AA^{\mathrm{T}}/A^{\mathrm{T}}A$ equal in decomposition

- AA^{T} and $A^{\mathrm{T}}A$ have equivalent eigen-decomposition
- We will prove these facts

1. AA^{T} and $A^{\mathrm{T}}A$ have the same rank

2.1 AA^{T} and $A^{\mathrm{T}}A$ have the same eigen-values

2.2 AA^{T} and $A^{\mathrm{T}}A$ have the same eigen-vectors (different for only up to an orthogonal transformation)

AA^{T} and $A^{\mathrm{T}}A$ have the same rank

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

$\because \operatorname{rank}(A) = \operatorname{rank}\left(A^{\mathrm{T}}\right)$, $\operatorname{rank}\left(A^{\mathrm{T}}A\right) = \operatorname{rank}\left(AA^{\mathrm{T}}\right)$

Proof $X^{\mathrm{T}}X = 0 \Rightarrow X = 0$

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

$AA^{\mathrm{T}}/A^{\mathrm{T}}A$ have 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 $AA^{\mathrm{T}} \leftrightarrow A^{\mathrm{T}}A$

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

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

AA^{T} and $A^{\mathrm{T}}A$ related through SVD

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

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

  $AA^{\mathrm{T}} = UD^2U^{\mathrm{T}}$

  (Proof omitted)

Gramian as kernel

- A kernel is a function that computes a distance $d(x_i, x_j)$ in a high dimensional mapped space ϕ without knowing $\phi(x_i)$ and $\phi(x_j)$
  - Good when $\phi(x)$ has way more features than x in the original space
    - But when number of datapoints is large, still better to compute $\phi(x_i)$ and $\phi(x_j)$
  - $d(x_i, x_j)$ often defined to be an inner product, $\langle\phi(x_i), \phi(x_j)\rangle$
    - More precisely, a Gramian