Matrix slicing

Relation to Covariance matrix computation and Low Rank models

S. Marchand-Maillet

Computer Science – University of Geneva

1 Introduction

Given two matrices $\mathbf{A} \in \mathbb{R}^{M\times N}, \mathbf{B} \in \mathbb{R}^{P\times N}$, we are interested in the product between $\mathbf{A}$ and the transpose of $\mathbf{B}$ ($\mathbf{A}\cdot\mathbf{B}^\top$). This product structure often arises naturally. For example (see below) with SVD : $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\cdot\mathbf{V}^\top$.

The standard matrix multiplication formula says that each element of the product matrix is an inner product as the product between the appropriate line from the first matrix and column from the second matrix.

For $\mathbf{U} \in \mathbb{R}^{M\times N}, \mathbf{V} \in \mathbb{R}^{N\times P}$: ([1]: we need column vectors to write the inner product)

$$(\mathbf{U}\cdot\mathbf{V})_{ij} = \sum_{k=1}^{N} \mathbf{U}_{ik}\mathbf{V}_{kj} \overset{[1]}{=} \sum_{k=1}^{N} (\mathbf{U}^\top)_{ki}\mathbf{V}_{kj} = \left\langle (\mathbf{U}^\top)_{:i}, \mathbf{V}_{:j} \right\rangle = ((\mathbf{U}^\top)_{:i})^\top \mathbf{V}_{:j} = \mathbf{U}_{i:}\mathbf{V}_{:j} \quad \forall\, i \in [\![M]\!], j \in [\![P]\!]$$

In our case ($\mathbf{U} \leftarrow \mathbf{A}$ and $\mathbf{V} \leftarrow \mathbf{B}^\top$ – see figure 1(left)):

$$(\mathbf{A}\cdot\mathbf{B}^\top)_{ij} = \left\langle (\mathbf{A}^\top)_{:i}, (\mathbf{B}^\top)_{:j} \right\rangle = \mathbf{A}_{i:}(\mathbf{B}^\top)_{:j} \quad \forall\, i \in [\![M]\!], j \in [\![P]\!] \tag{1}$$

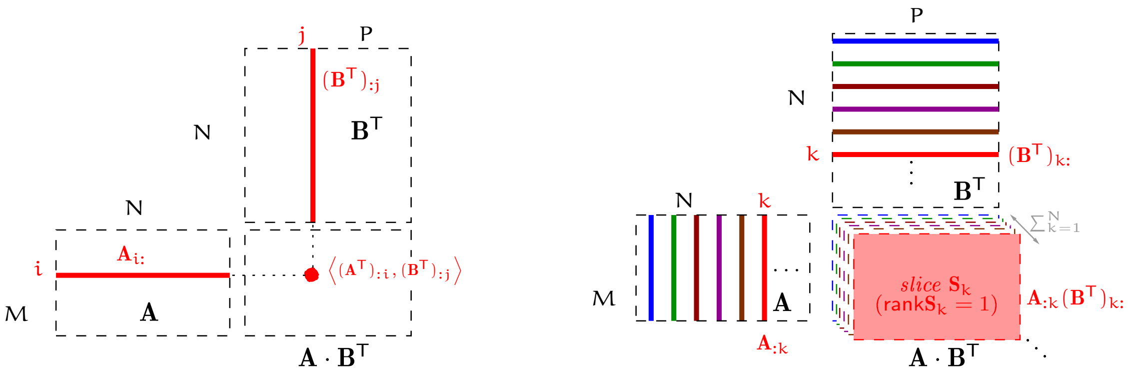

Figure 1: Two views of matrix multiplication. (left) *Standard view* where every element of the product is an inner product. (right) *Matrix slicing* where each slice is a 1-rank matrix produced from the duplication of columns of $\mathbf{A}$

However, an alternative reading (see figure 1(right)) given by the proposition below also shows useful.

Proposition 1 (Matrix slicing)**.** Given two matrices $\mathbf{A} \in \mathbb{R}^{M\times N}, \mathbf{B} \in \mathbb{R}^{P\times N}$, the product matrix $\mathbf{A}\cdot\mathbf{B}^\top \in \mathbb{R}^{M\times P}$ is such that

$$\mathbf{A}\cdot\mathbf{B}^\top = \sum_{k=1}^{N} \mathbf{A}_{:k}(\mathbf{B}^\top)_{k:}$$

Proof. Given $k \in [\![N]\!]$, call $\mathbf{C}_k \in \mathbb{R}^{M\times P}$ the matrix such that

$$(\mathbf{S}_k)_{ij} = \mathbf{A}_{ik}\mathbf{B}_{jk} \quad \forall\, i \in [\![M]\!], j \in [\![P]\!]$$

From (1), clearly

$$(\mathbf{A}\cdot\mathbf{B}^\top)_{ij} = \sum_{k=1}^{N} (\mathbf{S}_k)_{ij} \quad \Leftrightarrow \quad \mathbf{A}\cdot\mathbf{B}^\top = \sum_{k=1}^{N} \mathbf{S}_k \tag{2}$$

Now, if $(\mathbf{S}_k)_{ij} = \mathbf{A}_{ik}\mathbf{B}_{jk}$ then

$$\mathbf{S}_k = \mathbf{A}_{:k}(\mathbf{B}_{:k})^\top \quad \Rightarrow \quad \mathbf{S}_k = \underbrace{\mathbf{A}_{:k}}_{(M\times 1)} \underbrace{(\mathbf{B}^\top)_{k:}}_{(1\times P)}$$

As a result, from (2)

$$\mathbf{A}\cdot\mathbf{B}^\top = \sum_{k=1}^{N}\mathbf{S}_k = \sum_{k=1}^{N}\mathbf{A}_{:k}(\mathbf{B}^\top)_{k:}$$

□

Note. Note that $\mathbf{S}_k$ is a 1-rank matrix since all columns of $\mathbf{S}_k$ are multiples of the k^{th} column of $\mathbf{A}$

$$(\mathbf{S}_k)_{:i} = \mathbf{A}_{:k}(\mathbf{B}^\top)_{ki} = \mathbf{B}_{ik}\mathbf{A}_{:k}$$

Hence $\mathbf{A}\cdot\mathbf{B}^\top$ is a sum (superimposition) of 1-rank matrices $\mathbf{S}_k$ (slices). The rank of $\mathbf{A}\cdot\mathbf{B}^\top$ will therefore be determined (upper-bounded) by the rank of $\mathbf{A}$.

2 Application to covariance matrix computation

Given $\mathbf{X}\in\mathbb{R}^{D\times N}$ a matrix formed out N of D-dimensional samples. The covariance matrix $\mathbf{\Sigma}$ of data $\mathbf{X}$ is a $D\times D$ matrix such that

$$\mathbf{\Sigma}_{ij} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_{ik}-\overline{\mathbf{X}}_i)(\mathbf{X}_{jk}-\overline{\mathbf{X}}_j) \quad \forall\, i,j\in[\![D]\!] \tag{3}$$

where

$$\overline{\mathbf{X}} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}_{:k} = \frac{1}{N}\mathbf{X}\mathbf{1}_N \quad \text{and} \quad \mathbf{1}_N \quad \text{is the } N\text{-dimensional vector of all ones}$$

If we center the data, i.e we remove its mean to every line (feature), we get

$$\mathbf{X}\leftarrow\mathbf{X}-\overline{\mathbf{X}}\mathbf{1}_N^\top = \mathbf{X}\cdot(\mathbf{Id}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top) \quad \text{so that} \quad \overline{\mathbf{X}} = \frac{1}{N}\mathbf{X}\mathbf{1}_N = \mathbf{0}_D \qquad \text{(prove)}$$

Note. Matrix $\mathbf{C}_N = \mathbf{Id}_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N^\top$ is known as the centering matrix (of size N).

Exercise 1. Prove that $\mathbf{C}_N$ is symmetric and idempotent (i.e $\mathbf{C}_N^\top = \mathbf{C}_N$ and $\mathbf{C}_N\mathbf{C}_N = \mathbf{C}_N$). *Hint*: simply develop.

Assume that the data is centered, we can replace (3) by:

$$\mathbf{\Sigma}_{ij} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}_{ik}\mathbf{X}_{jk} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}_{ik}(\mathbf{X}^\top)_{kj}$$

Given $k\in[\![N]\!]$, let $\mathbf{S}_k\in\mathbb{R}^{D\times D}$ be such that $\mathbf{S}_k = \mathbf{X}_{:k}(\mathbf{X}^\top)_{k:}$. Then, clearly, $(\mathbf{S}_k)_{ij} = \mathbf{X}_{ik}(\mathbf{X}^\top)_{kj}$ So that,

$$\mathbf{\Sigma}_{ij} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{S}_k)_{ij} \quad \Leftrightarrow \quad \mathbf{\Sigma} = \frac{1}{N}\sum_{k=1}^{N}\mathbf{S}_k = \frac{1}{N}\sum_{k=1}^{N}\mathbf{X}_{:k}(\mathbf{X}^\top)_{k:}$$

By Proposition 1, for a centered matrix $\mathbf{X}$:

$$\mathbf{\Sigma} = \frac{1}{N}\mathbf{X}\mathbf{X}^\top$$

Note. The following properties can be derived:

1. Writing $\mathbf{\Sigma} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{S}_k$ (i.e that $\mathbf{\Sigma}$ is the average of slices $\mathbf{S}_k$) corresponds to the matrix version of the König-Huygens theorem saying that $\mathsf{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$. Since X is centered, $\mathbb{E}X = 0$ and we have $\mathsf{Var}(X) = \mathbb{E}[X^2]$, whose matrix-equivalent is
$$\mathbf{\Sigma} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{X}_{:k}(\mathbf{X}^\top)_{k:}$$
2. Should $\mathbf{X}$ not be centered, then
$$\mathbf{\Sigma} = \frac{1}{N}\mathbf{X}(\mathbf{Id} - \mathbf{1}\mathbf{1}^\top)(\mathbf{X}(\mathbf{Id} - \mathbf{1}\mathbf{1}^\top))^\top = \frac{1}{N}\mathbf{X}\mathbf{C}_N\mathbf{X}^\top$$

3 Application to Singular Value Decomposition and low rank models

For any matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$ there exist $\mathbf{U} \in \mathbb{R}^{M\times M}$, $\mathbf{\Sigma} \in \mathbb{R}^{M\times N}$ ("diagonal"), $\mathbf{V} \in \mathbb{R}^{N\times N}$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$.

Note. By definition:

1. $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and of rank M and N, respectively
2. If $M = N$, then we can consider that $\mathbf{V} = \mathbf{U}$ and obtain the classical eigensystem $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^\top$

Rewriting $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}.\mathbf{V}^\top$, Proposition 1 applies:
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}.\mathbf{V}^\top = \sum_{k=1}^{N} (\mathbf{U}\mathbf{\Sigma})_{:k}(\mathbf{V}^\top)_{k:}$$
Now, if $\underline{M \geqslant N}$, due to the structure of $\mathbf{\Sigma}$ (diagonal padded with zero-rows) we have
$$(\mathbf{U}\mathbf{\Sigma})_{:k} = \sigma_k \mathbf{U}_{:k} \quad \Rightarrow \quad (\mathbf{U}\mathbf{\Sigma})_{:k}(\mathbf{V}^\top)_{k:} = \sigma_k \mathbf{U}_{:k}(\mathbf{V}^\top)_{k:} \quad \text{for } k \in [\![N]\!]$$
Here again, slice $\mathbf{S}_k = \mathbf{U}_{:k}(\mathbf{V}^\top)_{k:}$ is a 1-rank matrix.
$$\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top = \sum_{k=1}^{N} \sigma_k \mathbf{U}_{:k}(\mathbf{V}^\top)_{k:} = \sum_{k=1}^{N} \sigma_k \mathbf{S}_k$$
Hence, the number of non-zero singular values (resp. eigen values if $M = N$) provides the rank of matrix $\mathbf{A}$.

If $\underline{M \leqslant N}$, columns $M+1, \cdots, N$ of $\mathbf{\Sigma}$ are zero-padding so that columns $M+1, \cdots, N$ of $\mathbf{U}\mathbf{\Sigma}$ will also be zero. We therefore similarly get the result that the number of non-zero singular values (resp. eigen values if $M = N$) provides the rank of matrix $\mathbf{A}$.

Note. The same applies considering that
$$\mathbf{A} = \mathbf{U} \cdot (\mathbf{V}\mathbf{\Sigma}^\top)^\top \qquad \text{(prove)}$$