Comparing Matrix Ranges

Karl Stratos

It is useful to view a matrix as representing a subspace. This is because it eliminates redundancy. For instance, if columns correspond to different features of data, adding a column that's a scalar multiple of an exiting one doesn't "improve" our data representation at all because the range of the matrix remains the same.

A key tool in comparing the ranges of a matrix pair is the concept of canonical angles between subspaces. It is mentally helpful to assume a setting where we have N samples of some object represented by d features in one view and d' features in another. We assume $N \geq \max(d, d')$ and organize these views as matrices $\boldsymbol{X} \in \mathbb{R}^{N \times d}$ and $\boldsymbol{Y} \in \mathbb{R}^{N \times d'}$. To be fully general, let $p = \mathrm{rank}\,(\boldsymbol{X})$ and $q = \mathrm{rank}\,(\boldsymbol{Y})$ with $p \geq q$.

1 Canonical Correlations Between Matrices

Following [1], for $i = 1 \ldots q$ we will write $\sigma_i(\boldsymbol{X}, \boldsymbol{Y})$ and call it the **i-th canonical correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$** to denote the cosine of the i-th canonical angle between $\mathrm{range}\,(\boldsymbol{X}), \mathrm{range}\,(\boldsymbol{Y}) \subset \mathbb{R}^N$. Recall that the i-th canonical angle is the smallest angle between any pair of nonzero vectors from $\mathrm{range}\,(\boldsymbol{X})$ and $\mathrm{range}\,(\boldsymbol{Y})$ under the constraint that they are view-wise orthogonal to vectors used to obtain the previous $i-1$ canonical angles. Because the range is simply all possible linear combination of columns, this problem can be posed as optimizing the column weights: for $i = 1 \ldots q$, find

$$(\bar{u}_i, \bar{v}_i) \in \underset{\substack{u \in \mathbb{R}^d:\ \boldsymbol{X}u \neq 0 \\ v \in \mathbb{R}^{d'}:\ \boldsymbol{Y}v \neq 0 \\ \langle \boldsymbol{X}u, \boldsymbol{X}u_j \rangle = \langle \boldsymbol{Y}v, \boldsymbol{Y}v_j \rangle = 0\ \forall j < i}}{\arg\max} \frac{\langle \boldsymbol{X}u, \boldsymbol{Y}v \rangle}{||\boldsymbol{X}u||\, ||\boldsymbol{Y}v||} \tag{1}$$

whereupon the i-th canonical correlation is obtained as

$$\sigma_i(\boldsymbol{X}, \boldsymbol{Y}) = \frac{\langle \boldsymbol{X}\bar{u}_i, \boldsymbol{Y}\bar{v}_i \rangle}{||\boldsymbol{X}\bar{u}_i||\, ||\boldsymbol{Y}\bar{v}_i||} \in [0, 1]$$

Note that because we are maximizing, $\sigma_i(\boldsymbol{X}, \boldsymbol{Y})$ corresponds to the cosine of an *accute* angle between a vector in $\mathrm{range}\,(\boldsymbol{X})$ and a vector in $\mathrm{range}\,(\boldsymbol{Y})$. Also, since the cosine is invariant to scaling, we can consider without loss of generality the following simpler objective equivalent to (1)

$$(\tilde{u}_i, \tilde{v}_i) \in \underset{\substack{u \in \mathbb{R}^d:\ ||\boldsymbol{X}u|| = 1 \\ v \in \mathbb{R}^{d'}:\ ||\boldsymbol{Y}v|| = 1 \\ \langle \boldsymbol{X}u, \boldsymbol{X}u_j \rangle = \langle \boldsymbol{Y}v, \boldsymbol{Y}v_j \rangle = 0\ \forall j < i}}{\arg\max} \langle \boldsymbol{X}u, \boldsymbol{Y}v \rangle \tag{2}$$

where we obtain $\sigma_i(\boldsymbol{X}, \boldsymbol{Y}) = \langle \boldsymbol{X}\tilde{u}_i, \boldsymbol{Y}\tilde{v}_i \rangle$.

As an exercise, it is useful to calculate the canonical correlations between $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^{3\times 2}$ whose ranges form two planes with angle $\pi/2$ in $\mathbb{R}^3$. They must intersect (why?). Verify that $\sigma_1(\boldsymbol{X}, \boldsymbol{Y}) = 1$ with solution unit vectors along the intersection and $\sigma_2(\boldsymbol{X}, \boldsymbol{Y}) \approx 0.707$ with solution unit vectors orthogonal to the intersection. The last canonical correlation can be viewed as a natural measure of difference between ranges.

The solution vectors themselves are interesting even though they are motivated simply as instruments for calculating the canonical correlations. We call $a_i = \boldsymbol{X}\tilde{u}_i$ and $b_i = \boldsymbol{Y}\tilde{v}_i$ the **i-th canonical vectors between $\boldsymbol{X}$ and $\boldsymbol{Y}$**. Their special property (besides being unit-length) is that $a_i^\top b_i = \sigma_i(\boldsymbol{X}, \boldsymbol{Y})$. In matrix form, we can organize $A = \boldsymbol{X}[\tilde{u}_1 \ldots \tilde{u}_q]$ and $B = \boldsymbol{Y}[\tilde{v}_1 \ldots \tilde{v}_q]$ which are now both $N \times q$ orthonormal matrices. Note that $\text{range}\,(B) = \text{range}\,(\boldsymbol{Y})$, so B is an orthonormal basis of $\text{range}\,(\boldsymbol{Y})$. But this is not just any orthonormal basis (e.g., obtained by running Gram-Schmidt on the columns of $\boldsymbol{Y}$)! It's particular orthonormal basis such that

$$A^\top B = \text{diag}\,(\sigma_1(\boldsymbol{X}, \boldsymbol{Y}) \ldots \sigma_q(\boldsymbol{X}, \boldsymbol{Y}))$$

It is easier to see the selectiveness of this subspace for $\text{range}\,(A) \subset \text{range}\,(\boldsymbol{X})$ when $p > q$. Suppose $\text{range}\,(\boldsymbol{X})$ is a plane and $\text{range}\,(\boldsymbol{Y})$ is a line forming angle $\pi/2$ in $\mathbb{R}^3$. In this case, there is only one canonical correlation with value $\sigma_1(\boldsymbol{X}, \boldsymbol{Y}) \approx 0.707$. The canonical vector b_1 spans the entire $\text{range}\,(\boldsymbol{Y})$, but the canonical vector a_1 spans a *specific* line in $\text{range}\,(\boldsymbol{X})$ that's closest to $\text{range}\,(\boldsymbol{Y})$.

In general, we can pick $m \leq q$ columns $A_m, B_m \in \mathbb{R}^{N\times m}$ of A and B corresponding to the top m canonical correlations $\sigma_1(\boldsymbol{X}, \boldsymbol{Y}) \geq \cdots \geq \sigma_m(\boldsymbol{X}, \boldsymbol{Y})$. Then A_m, B_m are orthonormal bases of m-dimensional subspaces of $\text{range}\,(\boldsymbol{X}), \text{range}\,(\boldsymbol{Y}) \subset \mathbb{R}^N$ that are constructed according to the greedy canonical correlation maximization process above. We will call these subspaces **rank-m best-match subspaces between** $\text{range}\,(\boldsymbol{X})$ **and** $\text{range}\,(\boldsymbol{Y})$.

2 How to Calculate Canonical Correlations

The q canonical correlations $\sigma_1(\boldsymbol{X}, \boldsymbol{Y}) \geq \cdots \geq \sigma_q(\boldsymbol{X}, \boldsymbol{Y})$ and the corresponding canonical vectors $A, B \in \mathbb{R}^{N\times q}$ can be calculated in a rather roundabout manner by first obtaining *some* orthonormal bases $R_{\boldsymbol{X}} \in \mathbb{R}^{N\times p}$ and $R_{\boldsymbol{Y}} \in \mathbb{R}^{N\times q}$ of $\text{range}\,(\boldsymbol{X})$ and $\text{range}\,(\boldsymbol{Y})$. How we obtain $R_{\boldsymbol{X}}, R_{\boldsymbol{Y}}$ is not important (Gram-Schmidt, SVD, QR decomposition, etc.). What *is* important is that because they are bases, their columns span all of the ranges. Thus we can consider the following problem equivalent to (2) (note the changed dimensions)

$$(u_i, v_i) \in \underset{\substack{u\in\mathbb{R}^p:\ ||R_{\boldsymbol{X}}u||=1 \\ v\in\mathbb{R}^q:\ ||R_{\boldsymbol{Y}}v||=1 \\ \langle R_{\boldsymbol{X}}u, R_{\boldsymbol{X}}u_j\rangle = \langle R_{\boldsymbol{Y}}v, R_{\boldsymbol{Y}}v_j\rangle = 0\ \forall j<i}}{\arg\max} \langle R_{\boldsymbol{X}}u, R_{\boldsymbol{Y}}v \rangle$$

and calculate $\sigma_i(\boldsymbol{X}, \boldsymbol{Y}) = \langle R_{\boldsymbol{X}} u_i, R_{\boldsymbol{Y}} v_i \rangle$. But $R_{\boldsymbol{X}}, R_{\boldsymbol{Y}}$ are moreover orthonormal, and this greatly simplifies the objective as

$$(u_i, v_i) \in \underset{\substack{u \in \mathbb{R}^p:\, ||u||=1 \\ v \in \mathbb{R}^q:\, ||v||=1 \\ u^\top u_j = v^\top v_j = 0\ \forall j < i}}{\arg\max} \quad u^\top R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}} v \tag{3}$$

Hence from (3) we see that $u_1 \ldots u_q \in \mathbb{R}^p$ and $v_1 \ldots v_q \in \mathbb{R}^q$ are left/right singular vectors of $R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}} \in \mathbb{R}^{p \times q}$ corresponding to the largest q singular values $\sigma_1(R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}}) \geq \cdots \geq \sigma_q(R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}})$. The i-th canonical correlation is given by

$$\sigma_i(\boldsymbol{X}, \boldsymbol{Y}) = \sigma_i(R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}})$$

and orthonormal bases of rank-m best-match subspaces are given by

$$A_m = R_{\boldsymbol{X}}[u_1 \ldots u_m] \in \mathbb{R}^{N \times m} \qquad B_m = R_{\boldsymbol{Y}}[v_1 \ldots v_m] \in \mathbb{R}^{N \times m}$$

3 Relation to Canonical Correlation Analysis

We assume that $\boldsymbol{X} \in \mathbb{R}^{N \times d}$ and $\boldsymbol{Y} \in \mathbb{R}^{N \times d'}$ are full-rank: they have dimensions d and d'. Let $\widetilde{\boldsymbol{X}} \in \mathbb{R}^{N \times d}$ and $\widetilde{\boldsymbol{Y}} \in \mathbb{R}^{N \times d'}$ denote the matrices after centering (i.e., we subtract the row average from every row). Since the matrices are full rank, a simple orthonormal basis of range $\left(\widetilde{\boldsymbol{X}}\right)$ is given by $R_{\boldsymbol{X}} = \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}})^{-1/2}$ (likewise for $\widetilde{\boldsymbol{Y}}$). Thus we can find orthonormal bases of rank-m best-match subspaces by computing the left $U_m \in \mathbb{R}^{d \times m}$ and right $V_m \in \mathbb{R}^{d' \times m}$ singular vectors of

$$R_{\boldsymbol{X}}^\top R_{\boldsymbol{Y}} = (\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}})^{-1/2} \widetilde{\boldsymbol{X}} \widetilde{\boldsymbol{Y}} (\widetilde{\boldsymbol{Y}}^\top \widetilde{\boldsymbol{Y}})^{-1/2} =: \boldsymbol{\Omega}$$

corresponding to singular values $\sigma_1(\boldsymbol{\Omega}) \geq \cdots \geq \sigma_m(\boldsymbol{\Omega})$. The i-th canonical correlation is $\sigma_i(\boldsymbol{\Omega})$ and the orthonormal bases are

$$A_m = \widetilde{\boldsymbol{X}}(\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}})^{-1/2} U_m \in \mathbb{R}^{N \times m} \qquad B_m = \widetilde{\boldsymbol{Y}}(\widetilde{\boldsymbol{Y}}^\top \widetilde{\boldsymbol{Y}})^{-1/2} V_m \in \mathbb{R}^{N \times m}$$

These orthonormal bases are precisely the m-dimensional linear transformation of data defined in CCA where we view rows of $\boldsymbol{X}, \boldsymbol{Y}$ as samples of random variables. Thus CCA is equivalent to finding rank-m best-match subspaces between the feature spans in two views, after centering.

References

[1] Golub, G. H. and Zha, H. (1994). Perturbation analysis of the canonical correlations of matrix pairs. *Linear Algebra and its Applications*, **210**, 3–28.