How to compare the size of covariance matrices?

2024-04-10
陳捷 Jesse C. Chen

Consider two covariance matrices $\mathbf{A}_{n\times n}$ and $\mathbf{B}_{n\times n}$. We say that $\mathbf{A}$ is **bigger** than $\mathbf{B}$, often denoted by $\mathbf{A} \geq \mathbf{B}$ or $\mathbf{A} \succeq \mathbf{B}$, if $\mathbf{A} - \mathbf{B}$ is semi-positive definite. Why do we use the "definiteness" of a matrix to compare the size of two covariance matrices?

First, notice that a covariance matrix is not only symmetrical, but also semi-positive definite. Consider a random vector $\boldsymbol{x} = (x_1, ..., x_n)^\top$. The covariance matrix is defined by

$$\mathbf{K} := \mathbf{E}[(\boldsymbol{x} - \mathbf{E}[\boldsymbol{x}])(\boldsymbol{x} - \mathbf{E}[\boldsymbol{x}])^\top].$$

Given any constant vector $\mathbf{v}$ of length n, we have

$$\mathbf{v}^\top \mathbf{K}\mathbf{v} = \mathbf{E}[\mathbf{v}^\top(\boldsymbol{x} - \mathbf{E}[\boldsymbol{x}])(\mathbf{v}^\top(\boldsymbol{x} - \mathbf{E}[\boldsymbol{x}]))^\top] \geq 0$$

by the definition of $\mathbf{K}$. Therefore, the covariance matrix $\mathbf{K}$ is semi-positive definite. In fact, $\mathbf{v}^\top\mathbf{K}\mathbf{v}$ is zero iff $\boldsymbol{x}$ has no variance at all.

There is another intuitive way of interpreting the definiteness described above. Consider the same vector $\mathbf{v}$ and the random vector $\boldsymbol{x}$. The dot product $\mathbf{v}^\top\boldsymbol{x}$ is the projection of the random vector from n-dimensional space on a one-dimensional space along the direction of $\mathbf{v}$, i.e., this collapse the n-dimensional random variable to a one-dimensional random variable through some linear combination. If we calculate the variance of the one-dimensional random variable $\mathbf{v}^\top\boldsymbol{x}$, we obtain

$$\begin{aligned}
\mathrm{Var}[\mathbf{v}^\top\boldsymbol{x}] &= \mathbf{E}[\mathbf{v}^\top\boldsymbol{x}(\mathbf{v}^\top\boldsymbol{x})^\top] - \mathbf{E}[\mathbf{v}^\top\boldsymbol{x}]\,\mathbf{E}[\mathbf{v}^\top\boldsymbol{x}]^\top \\
&= \mathbf{v}^\top\left(\mathbf{E}[\boldsymbol{x}\boldsymbol{x}^\top] - \mathbf{E}[\boldsymbol{x}]\,\mathbf{E}[\boldsymbol{x}]^\top\right)\mathbf{v} \\
&= \mathbf{v}^\top\mathbf{K}\mathbf{v}.
\end{aligned}$$

Notice that the variance assumes the exact form as before. And since variance is non-negative, it is clear that the covariance matrix must be semi-positive definite. That is, for any direction $\mathbf{v}$, the variance of "$\boldsymbol{x}$ projected on that direction" is (clearly) non-negative.

Motivated by the intuitive interpretation, lets now compare two covariance matrices. Let $\boldsymbol{x} = (x_1, ..., x_n)^\top$ and $\boldsymbol{y} = (y_1, ..., y_n)^\top$ be random vectors with mean $(0, ..., 0)^\top$ for simplicity. Let $\mathbf{A} = \mathbf{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ and $\mathbf{B} = \mathbf{E}[\boldsymbol{y}\boldsymbol{y}^\top]$ be the covariance matrices. Our goal is to compare $\mathbf{A}$ and $\mathbf{B}$ in some meaningful way. We can project $\boldsymbol{x}$ and $\boldsymbol{y}$ on a vector $\mathbf{v}$, and then compare the variance (non-negative real number) of the two projections. To make the comparison meaningful, it is reasonable to compare *all* possible projections, i.e., consider all possible choices of $\mathbf{v}$.

Formally, consider any vector $\mathbf{v}$. The projection of $\boldsymbol{x}$ on $\mathbf{v}$ is $\mathbf{v}^\top\boldsymbol{x}$. The variance of $\mathbf{v}^\top\boldsymbol{x}$ is

$$\begin{aligned}
\mathbf{E}[(\mathbf{v}^\top\boldsymbol{x})^2] &= \mathbf{E}[\mathbf{v}^\top\boldsymbol{x}\boldsymbol{x}^\top\mathbf{v}] \\
&= \mathbf{v}^\top\,\mathbf{E}[\boldsymbol{x}\boldsymbol{x}^\top]\mathbf{v} = \mathbf{v}^\top\mathbf{A}\mathbf{v}
\end{aligned}$$

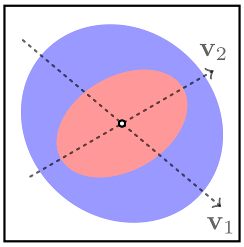
where $\mathbf{A}$ is the covariance matrix. Similarly, consider the same for $\boldsymbol{y}$. If we find that $\forall\mathbf{v}$,

$$\mathbf{v}^\top\mathbf{A}\mathbf{v} - \mathbf{v}^\top\mathbf{B}\mathbf{v} = \mathbf{v}^\top(\mathbf{A} - \mathbf{B})\mathbf{v} \geq 0,$$

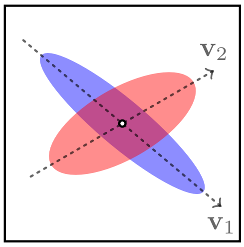
then, by definition, $\mathbf{A} - \mathbf{B}$ is semi-positive definite. Now we know why we say $\mathbf{A}$ is *larger* than $\mathbf{B}$ when $\mathbf{A} - \mathbf{B}$ is positive definite:

> If $\mathbf{A} - \mathbf{B}$ is positive definite, then *for all possible directions* $\mathbf{v}$, the variance of $\boldsymbol{x}$ is larger than $\boldsymbol{y}$'s. [1]

This interpretation of the partial ordering can be understood easily through visualisation. The following are representations of the distributions $\boldsymbol{x}$ and $\boldsymbol{y}$ where the two random vectors are two-dimensional:



case 1



case 2

Let $\boldsymbol{x}$ with covariance matrix $\mathbf{A}$ be the blue distribution and $\boldsymbol{y}$ with covariance matrix $\mathbf{B}$ be the red distribution. It is clear that in case 1, $\mathbf{A}$ is **bigger** than $\mathbf{B}$ since the variance of $\boldsymbol{x}$ is bigger that $\boldsymbol{y}$'s in *every* direction. (every possible direction of projection) However, the same statement is not true in case 2. In some directions (e.g. $\mathbf{v}_1$), the variance of $\boldsymbol{x}$ is larger; in other directions (e.g. $\mathbf{v}_2$), the variance of $\boldsymbol{y}$ is larger. Thus, $\mathbf{A}$ and $\mathbf{B}$ are not comparable by the partial order in case 2. ■

[1] This order of semi-positive definite matrices is called the Löwner ordering.