What Does Principal Component Analysis (PCA) Do?

date: 20260911
status: #finished
tags: #mathematics #linear_algebra #principal_component_analysis #eigenvector #eigenvalue

---

0. Introduction

The first part of the principal component analysis (PCA) is to obtain the set of eigenvectors $B \in \mathbb{R}^{k \times k}$ of the data matrix $X \in \mathbb{R}^{n \times k}$'s variance-covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$. This can be done by performing the SVD on the demeaned matrix $M_n X \in \mathbb{R}^{n \times k}$. This note aims at answering the question: what analysis we can do on B other than changing basis or doing dimension reduction.

---

1. Changing to the Orthonormal Eigenbasis

(i) Variance-Covariance Matrix After Changing Basis

Spectral theorem tells that we can do the decomposition on Σ:

$$\Sigma = B \Lambda B'$$

where B's columns are simply eigenvectors of Σ and $\Lambda \in \mathbb{R}^{k \times k}$ is a diagonal matrix whose diagonal entries are eigenvalues of Σ.

We can change the basis of X from standard basis to B via $XB \in \mathbb{R}^{n \times k}$.

Definition (PC Coordinate)

Changing basis is changing the coordinate system, and we name this coordinate system relative to B as "PC coordinate".

Definition (PC Factors)

Originally, the axes of the coordinate system are k number of features. After changing to PC coordinate, the axes become uninterpretable, and we name them as "PC factors".

Note that we don't necessarily have to utilize the full set of eigenvectors. Instead, we can simply pick a subset of eigenvectors with descending eigenvalues, i.e. picking eigenvectors that correspond to high eigenvalue, resulting in $B \in \mathbb{R}^{k \times j}$ where $j < k$. $XB \in \mathbb{R}^{n \times j}$ is known as the dimension reduction. For the following note we will use the term "changing coordinate" to avoid using "changing basis" as it implies that the dimension remains identical.

Besides, there are some properties regarding the variance after changing basis worth knowing.

Theorem

The variance-covariance matrix $\mathrm{Var}(XB)$ *after changing the coordinate is* Λ.

Proof

$$\begin{aligned}
\mathrm{Var}(XB) &= (XB)' M_n (XB) \\
&= B'(X' M_n X) B \\
&= B' \Sigma B \\
&= B' \Lambda B \\
&= \Lambda
\end{aligned}$$

where Λ is a diagonal matrix with entries being eigenvalues of Σ, corresponding to B.

The intuition of this theorem is that originally there is some correlation across axes (features) but after changing coordinate, axes are orthogonal.

(ii) Changing Coordinate Will Not Affect Total Variance

From the perspective of the total variance, after changing coordinate, it remains identical. Notice that total variance is computed by adding the variance across axes so it's computed by the trace operator.

Theorem

$$\operatorname{tr}(\Sigma) = \operatorname{tr}(\Lambda)$$

That is, the total variance after changing coordinate remains identical.

Proof

$$\begin{aligned}
\operatorname{tr}(\Sigma) &= \operatorname{tr}(B\Lambda B') \\
&= \operatorname{tr}\left(\begin{pmatrix} B_1 & \dots & B_k \end{pmatrix} \Lambda \begin{pmatrix} B_1' \\ \vdots \\ B_k' \end{pmatrix} \right) \\
&= \operatorname{tr}\left(\sum_i \Lambda_{ii} B_i B_i' \right) \\
&= \sum_i \Lambda_{ii} \operatorname{tr}(B_i B_i') \\
&= \sum_i \Lambda_{ii} \cdot 1 \\
&= \operatorname{tr}(\Lambda)
\end{aligned}$$

---

2. Variance Explained

(i) Total Variance Explained by Each Axis

Given the total variance of X, $\operatorname{tr}(\Sigma)$, and the variance-covariance matrix after changing coordinate, Λ, we can inspect

1. to what extent (in %) each axis explains the total variance of X (variance explained)
2. how fast do eigenvalues decay
3. how many components do you need to capture most of variance

by looking at:

$$\frac{1}{\operatorname{tr}(\Sigma)} \operatorname{Var}(XB) = \frac{1}{\operatorname{tr}(\Lambda)} \Lambda$$

(ii) In-Sample Variance Explained

In the above discussion, we utilize the data matrix X's variance-covariance matrix Σ to construct the PC coordinates, and samples after changing coordinate are obtained through XB where B stacks the eigenvectors of Σ horizontally. By spectrum decomposition, $\Sigma = B\Lambda B'$. Each eigenvector corresponds to a PC factor and the **variance explained** across different PC factors is measured by:

$$\frac{1}{\mathrm{tr}(\Lambda)}\Lambda.$$

This is called the **in-sample** variance explained.

(iii) Out-of-Sample Variance Explained

In financial domain, we often use the historical data X, computing the eigenbasis through spectral decomposition: $\Sigma = B\Lambda B'$. Then, we will change new data X^{out}'s coordinate using the eigenbasis derived from the historical data rather to obtain $X^{\text{out}}B$, and compute the **out-of-sample** variance explained:

$$\frac{1}{\mathrm{tr}(\Sigma^{\text{out}})} \ \mathrm{Var}(X^{\text{out}}B)$$

where Σ^{out} is the variance-covariance matrix of X^{out} and $\mathrm{tr}(\Sigma^{\text{out}})$ is its total variance.

Through the below derivation we can see that if the variance-covariance remains identical between historical and new data, i.e., $\Sigma^{\text{out}} = \Sigma$, the in-sample and out-of sample variance will remain identical as $\Lambda^{\text{out}} = \Lambda$ and $B^{\text{out}} = B$.

$$\begin{aligned}
\mathrm{Var}(X^{\text{out}}B) &= (X^{\text{out}}B)' M_n (X^{\text{out}}B) \\
&= B'\Sigma^{\text{out}}B \\
&= B'B^{\text{out}}\Lambda^{\text{out}}(B^{\text{out}})'B \\
&= \left[(B^{\text{out}})'B\right]'\Lambda^{\text{out}}\left[(B^{\text{out}})'B\right] \\
\Rightarrow \frac{1}{\mathrm{tr}(\Sigma^{\text{out}})} \ \mathrm{Var}(X^{\text{out}}B) &= \frac{1}{\mathrm{tr}(\Lambda^{\text{out}})}\left[(B^{\text{out}})'B\right]'\Lambda^{\text{out}}\left[(B^{\text{out}})'B\right].
\end{aligned}$$

By looking at the difference between the in-sample and out-of-sample variance explained, we can have some idea of to what extent the variance structure has changed. Specifically, for the in-sample variance explained, it's a diagonal matrix which means axes after changing coordinate are uncorrelated, however, it's not guaranteed that the out-of-sample variance explained has the same structure, and there might be covariance across axes. If the covariance is substantial, then the structure of variance across features has changed.

Remark

Nevertheless, one might ask why not just directly compare Σ and Σ^{out} ? The reason is that it might be easier to observe the difference using variance explained as the in-sample covariances after changing coordinate are uncorrelated and comparing zero with non-zero should be simpler.

3. Loadings and Scores

(i) Two Objects Produced by the Projection

After the changing to eigenbasis, using the PC coordinates, there are two matrices, and applied work gives each its own name.

Definition (Loading and Score)

Given the eigenbasis $B \in \mathbb{R}^{k \times j}$ and the projected data $XB \in \mathbb{R}^{n \times j}$:

- The **loading** b_{lm} is the entry of B in row l (a feature) and column m (a PC factor), representing the loading of feature l on PC m. Column m of B is the loading vector of the m-th PC factor.
- The **score** is an entry of XB. Row i of XB is sample i's coordinate in the new basis.

(ii) A Loading Is the Coefficient of a Linear Regression

Write $X_l \in \mathbb{R}^n$ for the l-th column of X (feature l across samples) and $Z_m = Xb_m \in \mathbb{R}^n$ for the m-th score column (PC factor m across samples).

Their sample variance-covariance $\widehat{\text{Cov}}[X_l, Z_m]$ is derived as follows.

$$
\begin{aligned}
\widehat{\text{Cov}}[X_l, Z_m] &= \frac{1}{n-1}(M_n X_l)'(M_n X b_m) \\
&= \frac{1}{n-1} X_l' M_n X b_m \\
&= e_l' \left[\frac{1}{n-1} X' M_n X \right] b_m \\
&= e_l' \Sigma b_m
\end{aligned}
$$

where e_l is the l-th standard basis vector, used with Xe_l to only pick out column l. Applying the eigen-equation $\Sigma b_m = \mu_m b_m$ where μ_m is the corresponding eigenvalue to b_m:

$$e_l' \Sigma b_m = e_l' \mu_m b_m = \mu_m b_{lm}$$

That is, the covariance between the old axis X_l and the new axis Z_m is exactly the loading b_{lm} scaled by the eigenvalue μ_m.

Moreover, we know that after changing coordinate, the eigenvalue μ_m is simply the sample variance of Z_m. Therefore,

$$\begin{aligned}
\widehat{\mathrm{Cov}}[X_l, Z_m] = \mu_m b_{l,m} &= \widehat{\mathrm{Var}}[Z_m] b_{l,m} \\
\Rightarrow b_{l,m} &= \frac{\widehat{\mathrm{Cov}}[X_l, Z_m]}{\widehat{\mathrm{Var}}[Z_m]}
\end{aligned}$$

Remark

$b_{l,m}$ is the regression coefficient when regressing X_l on Z_m.

The eigenvalue matrix Λ let us know the magnitude of the importance across PC factors and the eigenvector matrix B let us inspect how unknown PC factors influence known feature, demonstrating correlation among PC factors and features.

(iii) How to Compute the Correlation Among PC Factors and Features

Although the linear regression coefficients are intuitive, if one might prefer the correlation coefficients, we can computed them as follows.

$$\begin{aligned}
\mathrm{Corr}[X_l, Z_m] &= \frac{\widehat{\mathrm{Cov}}[X_l, Z_m]}{\widehat{\mathrm{SD}}[X_l]\widehat{\mathrm{SD}}[Z_m]} \\
&= \frac{\widehat{\mathrm{Var}}[Z_m] b_{l,m}}{\sqrt{\Sigma_{l,l}}\widehat{\mathrm{SD}}[Z_m]} \\
&= \left(\sqrt{\frac{\Lambda_{m,m}}{\Sigma_{l,l}}}\right) b_{l,m} \\
&= \left(\underbrace{(\mathrm{Diag}(\Sigma))^{-\frac{1}{2}} B \Lambda^{\frac{1}{2}}}_{\in \mathbb{R}^{k \times j}}\right)_{l,m}
\end{aligned}$$

$\mathrm{Diag}(\Sigma)$ is also a diagonal matrix whose diagonal entries are those of the Σ.

Moreover, $\mathrm{Diag}(\Sigma)$ can be computed through:

$$\sum_{i=1}^{k} E_i \Sigma E_i$$

where $E_i = e_i e_i' \in \mathbb{R}^{k\times k}$. The intuition is that ΣE_i extract the i-th column and $E_i(\Sigma E_i)$ extract the i-th entry of the i-th column. Therefore, in $E_i \Sigma E_i$, only the i-th row and i-column is non-zero, which is $\Sigma_{i,i}$, and all the other entries are 0.