Machine Learning and FinTech: PCA more

Huei-Wen Teng

National Yang Ming Chiao Tung University

A quick look at linear algebra and basis transformation

1. Standard vs. new basis

- Standard basis:

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

- New orthonormal basis (rotated 45°):

$$u_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad u_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

2. Points in the plane

$$P_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 4 \\ 0 \end{bmatrix}.$$

3. Coordinates in the new basis

Since $\{u_1, u_2\}$ is an **orthonormal basis**, the coordinates are simply the **inner products**:

$$c_1 = \langle u_1, P \rangle, \quad c_2 = \langle u_2, P \rangle.$$

- For P_1:

  $$\langle u_1, P_1 \rangle = \tfrac{1}{\sqrt{2}}(2+2) = 2\sqrt{2}, \quad \langle u_2, P_1 \rangle = \tfrac{1}{\sqrt{2}}(-2+2) = 0.$$

  So in the new basis:

  $$P_1 = (\, 2\sqrt{2},\, 0\,).$$

- For P_2:

  $$\langle u_1, P_2 \rangle = \tfrac{1}{\sqrt{2}}(4+0) = 2\sqrt{2}, \quad \langle u_2, P_2 \rangle = \tfrac{1}{\sqrt{2}}(-4+0) = -2\sqrt{2}.$$

  So in the new basis:

  $$P_2 = (\, 2\sqrt{2},\, -2\sqrt{2}\,).$$

4. Interpretation

- The **inner product with** u_1 gives the "shadow" of the point on the u_1-axis.
- The **inner product with** u_2 gives the "shadow" on the u_2-axis.
- Together, they are the new coordinates in this rotated orthonormal system.
- The **points in space remain the same**, only their coordinates change with the chosen basis.

Notations

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

The j-th feature

$$\mathbf{x}_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{pmatrix}$$

The i-th observation

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{in} \end{pmatrix}$$

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$$

$$\mathbf{X} = (x_i^\top, x_2^\top, \ldots, x_n^\top)$$

$^\top$: vector or matrix transpose

Notations

- The transpose of a vector $x_i^{\top} = (x_{i1}\ x_{i2}\ \cdots\ x_{ip})$
- The transpose of a matrix
- $$\mathbf{x}^{\top} = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ x_{1p} & x_{2p} & \cdots & x_{np} \end{pmatrix}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Preliminaries

- For a $p \times p$ matrix A, a non zero vector v, and a value λ.
- If $Av = \lambda v$, then v is an eigenvector of A with eigenvalue λ.

- If $\{v_1, \ldots, v_p\}$ is a basis, such that v_i is eigenvector of A with eigenvalue λ_i. Then,

$$A \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 & \cdots & \lambda_p v_p \end{pmatrix} = \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \cdots & \\ & & \lambda_p \end{pmatrix}$$

- Denote $V = \begin{pmatrix} v_1 & \cdots & v_p \end{pmatrix}$.

- We write $AV = V\Lambda$, or $A = V\Lambda V^{-1}$, and say A is diagonalizable.

The Principal Axis Theorem

- A matrix is symmatric if $A^\top = A$
- An orthonomal basis satisfies $V^{-1} = V^\top$
  - $VV^\top = VV^{-1} = I$
  - $v_i \perp v_j$ for $i \neq j$
  - $||v_i|| = 1$ for all i
- When A is a real symmetric matrix, then A admits an orthonormal eigen basis.

The Principal Axis Theorem

- When A is a real and symmetric matrix, A can be diagonalized, $A = V\Lambda V^\top$.
  - $\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \cdots & \\ & & \lambda_p \end{pmatrix}$ is the diagonal matrix with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$,
  - $V = \begin{pmatrix} v_1 \cdots v_p \end{pmatrix}$ is the orthonormal matrix of Eigen vectors $v_1, \ldots, v_p$, i.e., $V^\top V = I$.

Data matrix $\mathbf{X}$

- For an $n \times p$ data matirx $\mathbf{X}$, $\mathbf{X}^\top \mathbf{X}$ is real symmetric, because $(\mathbf{X}^\top \mathbf{X})^\top = \mathbf{X}^\top (\mathbf{X}^\top)^\top = \mathbf{X}^\top \mathbf{X}$
- Thus, $\mathbf{X}^\top \mathbf{X}$ is diagonalizable with orthonormal eigen basis.

PCA

- PCA actually seeks to find $w = (w_1, \ldots, w_p)^{\mathsf{T}}$ to maximize the variance of $Z_1 = \mathbf{X}w = X_1 w_1 + X_2 + w_2 + \cdots X_p w_p$.
- PAC aims at $\max_{\|w\|^2=1} \text{var}(Z_1)$
- Because Z_1 has a column mean 0, its variance equals

$$Var(Z_1) = \frac{1}{n}\|Z_1\|^2 = \frac{1}{n}\|\mathbf{X}w\|^2$$

- PCA equals to $\max_{\|w\|^2=1} \|\mathbf{X}w\|^2$

- However

$$
\begin{aligned}
\|\mathbf{X}w\|^2 &= w^\top \mathbf{X}^\top \mathbf{X} w = w^\top V \Lambda V^\top w \\
&= \tilde{w}^T \begin{pmatrix} \lambda_1 & & \\ & \vdots & \\ & & \lambda_p \end{pmatrix} \tilde{w} \\
&= \lambda_1 \tilde{w}_1^2 + \cdots + \lambda_p \tilde{w}_p^2 \\
&\leq \lambda_1
\end{aligned}
$$

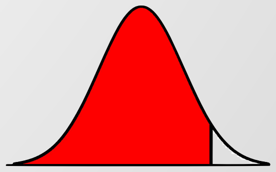
$$
V^\top w = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} w = \tilde{w}
$$

How do we do eigen decomposition for $\mathbf{X}^\top\mathbf{X}$?

- The eigen decomposition procedure

$(\mathbf{X}^\top\mathbf{X}) = V\Lambda V^{-1}$, where V is the matrix of eigenvectors, and Λ is the diagonal matrix with eigenvalues in the diagonal.

- Singular Value Decomposition (SVD). See next.



Singular Value Decomposition (SVD) I

The SVD of the $N \times p$ matrix X has the form $X = UDV'$.

- $U = (u_1, \ldots, u_N)$ is an $N \times N$ orthogonal matrix. $\{u_1, \ldots, u_N\}$ form an orthonormal basis for the space spanned by the column vectors of X.
- $V = (v_1, \ldots, v_p)$ is an $p \times p$ orthogonal matrix. $\{v_1, \ldots, v_p\}$ form an orthonormal basis for the space spanned by the raw vector of X.
- D is an $N \times p$ rectangular matrix with nonzero elements along the first $p \times p$ submatrix diagonal. $diag(d_1, d_2, \ldots, d_p)$, $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$ are the singular values of X with $N > p$. Let plum denote matrix and vector transpose.

Singular Value Decomposition (SVD) II

With SVD of X, we have

$$
\begin{aligned}
X'X &= (UDV')'(UDV') \\
&= VD'U'UDV' \\
&= VD'DV' \\
&= VD^2V'.
\end{aligned}
$$

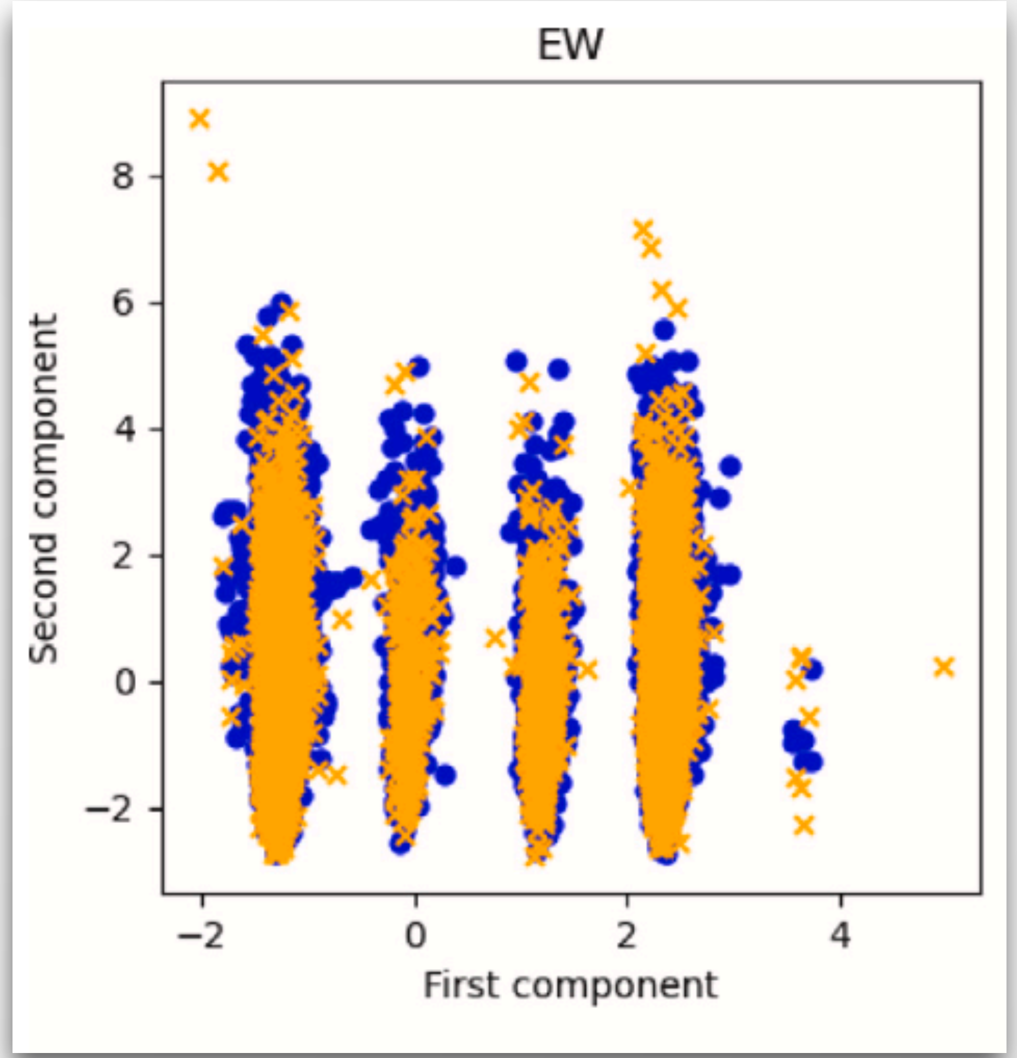
Here, $D^2 = D'D$. If you have the SVP, you already have the Eigen value decomposition for $X'X$.

- The columns of V (i.e., v_j, $j = 1, \ldots, p$) are the eigenvectors of $X'X$. They are called *principle component direction* of X.
- The diagonal values in D (i.e., d_1, $j = 1, \ldots, p$) are the square roots of the eigenvalues of $X'X$.

What do we use PCA for?

- Dimension reduction for visualization
- Insights? Rescaled Cluster-then-predict

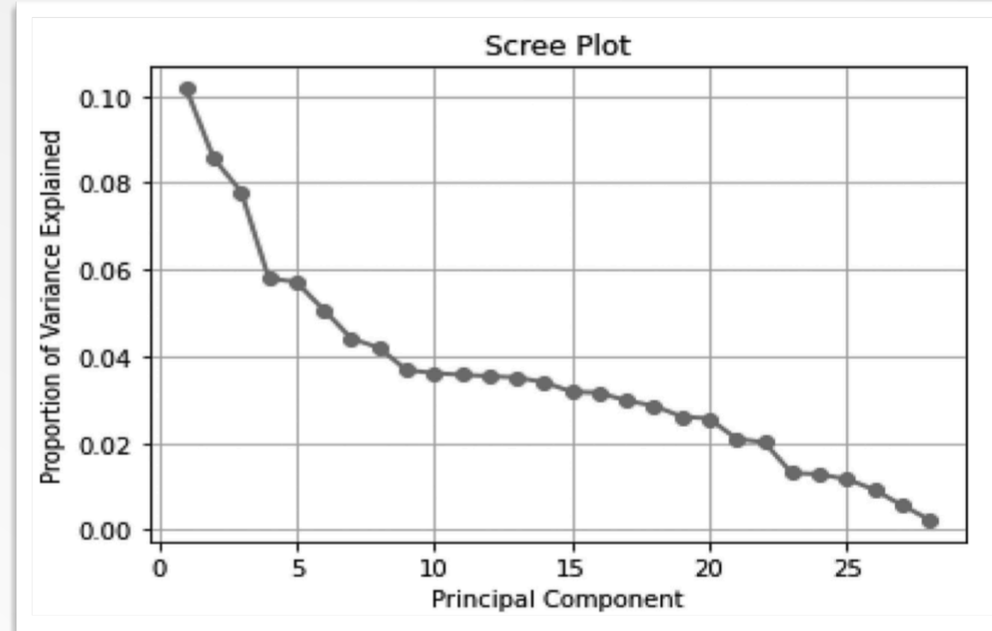
Teng et al (2024)



https://www.sciencedirect.com/science/article/abs/pii/S1057521923005215

What do we use PCA for?

- Feature engineering
  - Find representative feature
  - avoid multi-collinearity problems
  - improve prediction



Tuan et al (2023)

Table 3. Comparison of bank failure prediction methods: unadjusted data and after PCA.

Panel A: Unadjusted data				
In-sample test	Logistic regression	KNN	Decision Tree	Random Forest
Accuracy	0.6325	0.9459	1.0000	1.0000
Precision	0.6261	0.9422	1.0000	1.0000
Recall	0.4625	0.9378	1.0000	1.0000
Out-of-sample test				
Accuracy	0.6125	0.5689	0.5111	0.7076
Precision	0.2109	0.1847	0.1574	0.2389
Recall	0.5482	0.5227	0.4961	0.4070
Panel B: After PCA				
In-sample test				
Accuracy	0.6169	0.8578	1.0000	1.0000
Precision	0.6154	0.8544	1.0000	1.0000
Recall	0.4050	0.8260	1.0000	1.0000
Out-of-sample test				
Accuracy	0.5154	0.8109	0.7005	0.8057
Precision	0.1563	0.4372	0.2964	0.4278
Recall	0.4845	0.7735	0.6805	0.7558

https://www.tandfonline.com/doi/full/10.1080/13504851.2023.2186353

Probabilistic view using Eigen decomposition (Skipped)

- Suppose $\chi \sim N_p(\mathbf{0}, \Sigma)$. Find $w = (w_1, \ldots, w_p)^\top$ satisfying $\|w\|^2 = 1$ to maximize Var$(w^\top \chi)$

$$
\begin{aligned}
\mathrm{Var}(w^\top \chi) &= w^\top \Sigma w \\
&\approx \frac{1}{n} w^\top X^\top X w = \frac{1}{n} w^\top V \Lambda V^\top w \\
&= \frac{1}{n} \tilde{w}^\top \Lambda \tilde{w} \\
&= \frac{1}{n} (\lambda_1 \tilde{w}_1^2 + \cdots + \lambda_p \tilde{w}_p^2) \\
&\leq \frac{1}{n} \lambda_1
\end{aligned}
$$

Machine Learning and FinTech: PCA more

Huei-Wen Teng

National Yang Ming Chiao Tung University