Stat 6341.001: Numerical Linear Algebra and Statistical Computing, Fall 2026 Topic 13

Principal Component Analysis (PCA)

Lecturer: Nan Wu *Scribe: Chuqi Wang*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

1 Introduction

We have a high dim data set. It distributes around a low dim set.
Dimension Reduction: Reorganize the data in a low dim space. We want to preserve some properties of the data set in the low dim representation.
Principal Component Analysis (PCA).

1. Algorithm.
2. In what sense it reorganizes the data.
3. Efficient algorithm through SVD.

PCA is a linear dimension reduction method.
Data set:

$$\{X_1, \dots, X_n\} \subset \mathbb{R}^p, \qquad p \text{ is large.}$$

$$X_i = \begin{bmatrix} \vdots \\ \vdots \end{bmatrix}, \qquad p \text{ vector.}$$

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i \in \mathbb{R}^p.$$

2 Sample Covariance Matrix

$$C_n = \sum_{i=1}^{n} \underbrace{(X_i - \mu_n)}_{\mathbb{R}^{p\times 1}} \underbrace{(X_i - \mu_n)^T}_{\mathbb{R}^{1\times p}} \in \mathbb{R}^{p\times p}. \tag{*}$$

$$\Sigma_n = \frac{1}{n-1} C_n \qquad \text{Sample covariance matrix.}$$

Properties of C_n.

1. C_n is symmetric.
2. $$C_n = U\Lambda U^T, \qquad U \in O(p).$$

   U is an orthogonal $p \times p$ matrix, i.e.

$$UU^T = U^T U = I_{p\times p}.$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}, \qquad U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_p \\ | & & | \end{bmatrix}.$$

$$\lambda_1 \geq \cdots \geq \lambda_p \geq 0.$$

$\{u_1, \ldots, u_p\}$ are orthonormal eigenvectors, i.e. C_n is positive semi-definite. This follows from the structure of C_n in $(*)$ (e.g. eigenvalues ≥ 0).

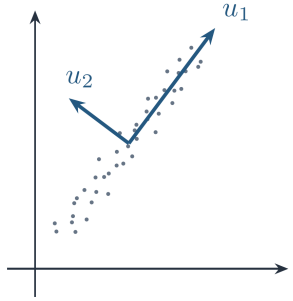
$(\lambda_i, u_i)_{i=1}^p$ are the orthonormal eigenpairs of C_n.

2.1 Understand Σ_n (or C_n)

u_i : direction of the ith largest variance of the data $\{X_1, \ldots, X_n\}$.

$\underbrace{\frac{\lambda_i}{n-1}}_{\text{eigenvalue of } \Sigma_n}$: ith largest variance.

$p = 2, \qquad \{X_1, \ldots, X_n\} \subset \mathbb{R}^2$



3 PCA Algorithm

Input: $\{X_1, \ldots, X_n\} \subset \mathbb{R}^p$.
# of principal components: k.
Step 1. Construct C_n using $\{X_1, \ldots, X_n\}$.
Step 2. Find the eigendecomp of C_n:

$$C_n = U \Lambda U^T.$$

$$U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_p \\ | & & | \end{bmatrix}, \qquad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_p \end{bmatrix}, \qquad \lambda_1 \geq \cdots \geq \lambda_p \geq 0.$$

Step 3. Project $X_i - \mu_n$ onto the first k eigenvectors.

$$\widetilde{U} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_k \\ | & & | \end{bmatrix}.$$

Projection:

$$\widetilde{U}^T (X_i - \mu_n) = \widetilde{X}_i \in \mathbb{R}^k.$$

$$X_i \xrightarrow{\text{PCA}} \widetilde{X}_i = \begin{bmatrix} u_1^T (X_i - \mu_n) \\ \vdots \\ u_k^T (X_i - \mu_k) \end{bmatrix} \in \mathbb{R}^k.$$

4 In What Sense PCA Reorganizes the Data

PCA reduce the dimension of the data from $\mathbb{R}^p$ to $\mathbb{R}^k$.

$$\sum_{i=1}^{n} \widetilde{X}_i = \widetilde{U}^T \left(\underbrace{\sum_{i=1}^{n} (X_i - \mu_n)}_{=0} \right) = 0.$$

PCA reorganize the data in the following sense.

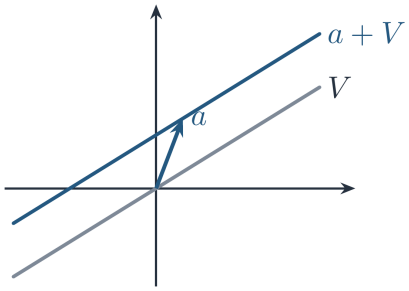
1. **Projection with maximum variability.** Project $\{X_1 - \mu_n, \ldots, X_n - \mu_n\}$ onto the principal components corresponding to the directions with k largest variances of the data.
2. It finds a k-dim affine subspace of $\mathbb{R}^p$ that best fits the data in the least square sense.

4.1 Recall Affine Space

V is a vector subspace of $\mathbb{R}^p$, $a \in \mathbb{R}^p$.

$$\text{affine subspace} \qquad a + V = \{a + v \mid v \in V\}.$$

$$\dim(a + V) = \dim(V).$$



$$\dim(V) = k, \qquad V \text{ is a subspace of } \mathbb{R}^p.$$

$\{\widetilde{u}_1, \ldots, \widetilde{u}_k\}$ form an orthonormal basis of V.

$$\beta_i \in \mathbb{R}^k, \qquad i = 1, \ldots, n.$$

$$\beta_i = \begin{bmatrix} \beta_{i,1} \\ \vdots \\ \beta_{i,k} \end{bmatrix} \in \mathbb{R}^k, \qquad \widetilde{U} = \begin{bmatrix} | & & | \\ \widetilde{u}_1 & \cdots & \widetilde{u}_k \\ | & & | \end{bmatrix}.$$

$$y_i = a + \widetilde{U}\beta_i = a + \sum_{j=1}^{k} \beta_{i,j} \widetilde{u}_j \in a + V.$$

4.2 Least Square Metric

Given $\{X_1, \dots, X_n\} \subset \mathbb{R}^p$ and an affine subspace $a + V$ of $\mathbb{R}^p$,

$$\{y_1, \dots, y_n\} \subset a + V.$$

$$\widetilde{U} = \begin{bmatrix} | & & | \\ \widetilde{u}_1 & \cdots & \widetilde{u}_k \\ | & & | \end{bmatrix}.$$

$\{\widetilde{u}_i\}$ are orthonormal basis of V. We can find $\beta_1, \dots, \beta_n \in \mathbb{R}^k$ s.t.

$$y_i = a + \widetilde{U}\beta_i.$$

$$I\Big(\underbrace{a, \widetilde{U}}_{\substack{\text{determines the affine} \\ \text{subspace } a+V}} , \underbrace{\beta_1, \dots, \beta_n}_{\text{determines } y_i \in a+V} \Big) = \sum_{i=1}^n \|X_i - y_i\|_{\mathbb{R}^p}^2$$

$$= \boxed{\sum_{i=1}^n \|X_i - a - \widetilde{U}\beta_i\|_{\mathbb{R}^p}^2}.$$

4.3 PCA Minimizes the Least Square Metric

Theorem 4.1. *Given $\{X_1, \dots, X_n\} \subset \mathbb{R}^p$, let*

$$a = \mu_n = \frac{1}{n}\sum_{i=1}^n X_i \in \mathbb{R}^p, \qquad \widetilde{U} = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_k \\ | & & | \end{bmatrix} \in \mathbb{R}^{p \times k}.$$

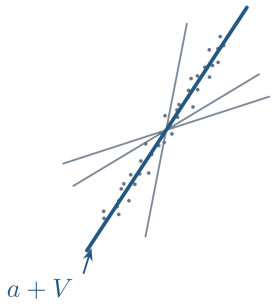
$u_1, \dots, u_k$ are the orthonormal eigenvectors of C_n corresponding to the decreasing order of the eigenvalues.

$$\beta_i = \widetilde{X}_i \qquad \textit{(output of PCA).}$$

Then $I(a, \widetilde{U}, \beta_1, \dots, \beta_n)$ is minimized.

I.e. PCA finds $y_1, \dots, y_n$ in the affine space $\mu_n + V$ with coordinates $\beta_1, \dots, \beta_n$ and V generated by $u_1, \dots, u_k$ so that

$$\sum_{i=1}^n \|X_i - y_i\|^2 \text{ is minimized.}$$



5 Complexity of the PCA Algorithm

Step 1.

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad C_n = \sum_{i=1}^{n}(X_i - \mu_n)(X_i - \mu_n)^T.$$

Recall the Big O notation. Complexity (# of operations, e.g. addition and multiplication).

Complexity in step 1:

$$\underbrace{n}_{\mu_n} + \underbrace{p \cdot p \cdot n}_{C_n} \leq 2p^2 n = O(p^2 n).$$

Step 2. Eigendecomposition. Complexity $O(p^3)$.

Step 3. Projection. Complexity is dominated by the complexity in the first 2 steps.

Total complexity is

$$O(p^2 n + p^3).$$

6 Singular Value Decomposition (SVD)

$$M \in \mathbb{R}^{p\times n}.$$

$$\text{SVD:} \qquad M = USV^T.$$

$$\begin{aligned} U:&\quad p \times p \text{ orthogonal matrix,}\\ V:&\quad n \times n \text{ orthogonal matrix,}\\ &S \in \mathbb{R}^{p\times n}. \end{aligned}$$

$$S = \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_{\min(p,n)} \\ & 0 & \end{bmatrix}}_{p\geq n} \quad \text{or} \quad \underbrace{\begin{bmatrix} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_{\min(p,n)} & \end{bmatrix}}_{p\leq n}.$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(p,n)} \geq 0.$$

They are singular values of M. Column vectors of U (or V) are the left (right) singular vectors.

$$Y = \begin{bmatrix} | & & | \\ X_1 - \mu_n & \cdots & X_n - \mu_n \\ | & & | \end{bmatrix} \in \mathbb{R}^{p\times n}.$$

$$\text{Observation:} \qquad \boxed{C_n = YY^T}.$$

SVD of Y:

$$Y = \widetilde{U}\Sigma\widetilde{V}^T.$$

$$\begin{aligned} C_n &= \widetilde{U}\Sigma\underbrace{\widetilde{V}^T V}_{I_{n\times n}}\Sigma^T U^T \\ &= \widetilde{U}\Sigma\Sigma^T\widetilde{U}^T \\ &= U\Lambda U^T \qquad \text{eigendecomposition.} \end{aligned}$$

$$\implies \qquad \underbrace{\lambda_i}_{\text{eigenvalues of } C_n} = \sigma_i^2 \geq 0, \qquad \Lambda = \Sigma\Sigma^T, \qquad U = \widetilde{U}.$$

First k singular vectors of Y can be calculated in $O(pnk)$.

$$pnk \leq p^2 n + p^3.$$

Use SVD of Y to replace the first two steps of PCA.