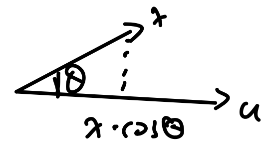
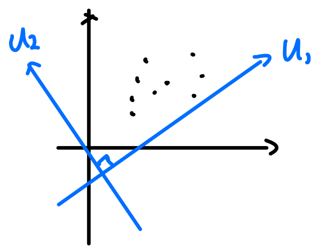
PCA Reviews
- 2 — Prospectives of PCA
- PCA Drawbacks



1) Prospective 1 (maximum projection variance)



$$\begin{cases} \underset{u}{\arg\max}\ \frac{1}{N}\sum_{i=1}^{N}(u^T x_i)^2 \\ \text{subject to } u^T u = 1 \end{cases}$$

$(x^T u)u$

$$\begin{cases} \underset{u}{\arg\max}\ u^T \Sigma u, \text{ where } \Sigma = \frac{1}{N}\sum_{i=1}^{n} x_i x_i^T \\ \text{subject to } u^T u = 1 \end{cases}$$

$*$

u is an eigenvector of Σ

λ is an eigenvalue of Σ

Summary.

① Centering Data

② Find the covariance matrix $\Sigma = x^T x$, $x = \begin{pmatrix} \cdots x_1 \cdots \\ \cdots x_2 \cdots \\ \vdots \\ \cdots x_n \cdots \end{pmatrix}_{n \times D}$

③ Find eigenvalues and eigenvectors of Σ

2) Prospective 2 (minimum Reconstructive cost)

orthogonal basis $U = [u_1\ u_2 \cdots u_d]$

$$x_i = \sum_{j=1}^{d}(u_j^T x_i)u_j = \sum_{j=1}^{d} u_j u_j^T x_i = UU^T x_i \quad (\text{real } x_i)$$

$$\hat{x}_i = \sum_{j=1}^{M}(u_j^T x_i)u_j = UU^T x_i \quad (\text{estimated } x_i, \text{ denoted as } \hat{x})$$

Reconstruction Error formula : $E = \frac{1}{N}\sum_{i=1}^{N}\|x_i - \hat{x}_i\|^2$

$$E = \frac{1}{N}\sum_{i=1}^{N}\left\|\sum_{j=1}^{d} u_j u_j^T x - \sum_{j=1}^{M} u_j u_j^T x\right\|^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left\|\left(\sum_{j=M+1}^{d} u_j u_j^T\right) x\right\|^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{j=M+1}^{d} u_j u_j^T\right) x\right]^T\left[\left(\sum_{k=M+1}^{d} u_j u_j^T\right) x\right]$$

$$= \frac{1}{N}\sum_{i=1}^{n} x_i^T\left(\sum_{j=M+1}^{d} u_j u_j^T\right)\left(\sum_{k=M+1}^{d} u_j u_j^T\right) x$$

$$= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=M+1}^{d}(x_i^T u_j)(u_j^T x_i)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=M+1}^{d} (u_j^T x_i)(x_i^T u_j)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{j=M+1}^{d} u_j^T (x_i x_i^T) u_j$$

$$= \sum_{j=M+1}^{d} u_j^T \left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T\right) u_j$$

$$= \sum_{j=M+1}^{d} u_j^T \Sigma u_j$$ * minimize this

Supplement

$$\tilde{V} = (V_1 \ V_2 \ \cdots \ V_k) \qquad k<d$$

$$\hat{x}_i = (x_i^T V_1) V_1 + (x_i^T V_2) V_2 + \cdots + (x_i^T V_k) V_k$$

$$= V_1 V_1^T x_i + V_2 V_2^T x_i + \cdots V_k V_k^T x_i$$

$$= (V_k V_k^T) x_i$$

$$= \tilde{V} \tilde{V}^T x_i$$

In paper, formula (4). $\boxed{\min_{V_k} \sum_{i=1}^{n} \| x_i - \tilde{V} \tilde{V}^T x_i \|^2}$

Sparse PCA

Algorithm 1. General SPCA Algorithm

1. Let A start at V[, 1 : k], the loadings of the first k ordinary principal components.
2. Given a fixed $\mathbf{A} = [\alpha_1, \ldots, \alpha_k]$, solve the following elastic net problem for $j = 1, 2, \ldots, k$

$$\beta_j = \arg\min_{\beta} (\alpha_j - \beta)^T \mathbf{X}^T \mathbf{X} (\alpha_j - \beta) + \lambda \|\beta\|^2 + \lambda_{1,j} \|\beta\|_1$$

3. For a fixed $\mathbf{B} = [\beta_1, \cdots, \beta_k]$, compute the SVD of $\mathbf{X}^T\mathbf{X}\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, then update $\mathbf{A} = \mathbf{U}\mathbf{V}^T$.
4. Repeat Steps 2–3, until convergence.
5. Normalization: $\hat{V}_j = \frac{\beta_j}{\|\beta_j\|}$, $j = 1, \ldots, k$.

Why sparse : L_1 penalty

how to solve : PCA → regression problem

Some Remarks of Paper's theorem:

①

$$\begin{cases} \arg\min \sum_{i=1}^{n} \| x_i - a a^T x_i \|^2 \\ \text{subject to } \|a\|^2 = 1 \end{cases}$$

let $\beta = a$,

$$\begin{cases} \arg\min \sum_{i=1}^{n} \| x_i - a\beta^T x_i \|^2 \\ \text{subject to } \|a\|^2 = 1, \ a = \beta \end{cases}$$

· Theorem 1: (first principal component)

$$(\hat{a}, \hat{\beta}) = \arg\min_{a, \beta} \sum_{i=1}^{n} \| x_i - a\beta^T x_i \|^2 + \lambda_0 \|\beta\|^2 \quad \triangle$$

$$\text{Subject to } \|a\|^2 = 1$$

· SPCA criterion (add lasso term and form a Elastic Net)

$$(\hat{a}, \hat{\beta}) = \begin{cases} \arg\min_{a, \beta} \sum_{i=1}^{n} \| x_i - a\beta^T x_i \|^2 + \lambda_0 \|\beta\|^2 + \lambda_1 \|\beta\|_1 \\ \text{Subject to } \|a\|^2 = 1 \end{cases}$$

· Theorem 2: (first k principal component)

$$A_{p\times k} = [a_1, a_2, \cdots, a_k]$$

$$B_{p\times k} = [\beta_1, \beta_2, \cdots, \beta_k]$$

$$(A, B) = \begin{cases} \arg\min_{A, B} \sum_{i=1}^{n} \| x_i - AB^T x_i \|^2 + \lambda_0 \sum_{j=1}^{k} \|\beta_j\|^2 \\ \text{subject to } A^T A = I_{k\times k} \end{cases}$$

· SPCA criterion in n-d

$$(A, B) = \begin{cases} \arg\min_{A, B} \sum_{i=1}^{n} \| x_i - AB^T x_i \|^2 + \lambda_0 \sum_{j=1}^{k} \|\beta_j\|^2 + \sum_{j=1}^{k} \lambda_{1j} \|B_j\|_1 \\ \text{subject to } A^T A = I_{k\times k} \end{cases} \quad ☆☆☆$$

Two problems:

1. how to solve this optimization problem? → Fix one variable
2. $\|X - AB^T X\|^2$ is hard to optimize! → Change to regression problem $\|Y - X^T\beta\|^2$

Numerical Solution.

B Given A:

$$\sum_{i=1}^{n} \|x_i - AB^T x_i\|^2 = \|X - AB^T X\|^2 \qquad A = B$$

$$= \|X - XBA^T\|^2$$

let $A_\perp$ be any orthonormal matrix such that $[\underset{p\times k}{A} ; \underset{k\times p}{A_\perp}]$ is $p\times p$ orthonormal matrix.

Project $\|X - XBA^T\|^2$ to A and $A_\perp$:

$$
\begin{aligned}
&\|X - XBA^T\|^2 \\
&= \|(X - XBA^T)\cdot A_\perp\|^2 + \|(X - XBA^T)A\|^2 \\
&= \| XA_\perp - XBA^TA_\perp\|^2 + \|XA - XBA^TA\|^2 \qquad \begin{cases} A^TA = I \\ A^TA_\perp = 0 \end{cases} \\
&= \|XA_\perp\|^2 + \|XA - XB\|^2 \\
&= \|XA_\perp\|^2 + \sum_{j=1}^{k} \|X\alpha_j - X\beta_j\|^2
\end{aligned}
$$

let $\underline{Y = X\alpha_j}$

$$\Rightarrow \|XA_\perp\|^2 + \sum_{j=1}^{k} \|Y_j - X\beta_j\|^2 \qquad (A_\perp \text{ is given})$$

Hence:

$$
\begin{aligned}
\beta_j &= \arg\min_{\beta_j} \|Y_j^* - X\beta_j\|^2 + \lambda_0\|\beta_j\|^2 + \lambda_{1j}\|\beta_j\|_1 \\
&= \arg\min_{\beta_j} (X\alpha_j - X\beta_j)^T(X\alpha_j - X\beta_j) + \lambda_0\|\beta_j\|^2 + \lambda_{1j}\|\beta_j\|_1 \\
&= \arg\min_{\beta_j} \left[(\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) \right] + \lambda_0\|\beta_j\|^2 + \lambda_{1j}\|\beta_j\|_1
\end{aligned}
$$

$$\Rightarrow \boxed{\begin{cases} \arg\min_{\beta_j} \left[(\alpha_j - \beta_j)^T X^T X (\alpha_j - \beta_j) \right] + \lambda_0\|\beta_j\|^2 + \lambda_{1j}\|\beta_j\|_1 \\ \text{Subject to } A^TA = I \end{cases} \quad \text{Algorithm step 2}}$$

A Given B: (Algorithm step 3!)

SPCA criterion copy:

$$(A, B) = \begin{cases} \arg\min_{A,B} \sum_{i=1}^{n} \|x_i - AB^Tx_i\|^2 + \underbrace{\lambda_0 \sum_{j=1}^{k} \|\beta_j\|^2}_{\times} + \underbrace{\sum_{j=1}^{k} \lambda_{1j}\|\beta_j\|_1}_{\times} \\ \text{subject to } A^TA = I_{k\times k} \end{cases}$$

If B is fixed. we can ignore the penalty part

$$\begin{cases} \arg\min_{A} \sum_{i=1}^{n} \|\underline{X_i - XBA^T}\|^2 \\ \text{Subject to } A^TA = I_{k\times k} \end{cases}$$

$$\Rightarrow \underline{(X^TX)B = U\Sigma V^T} \text{ and set } \hat{A} = UV^T \qquad *$$

Supplement:

Reduced Rank Procrustes Rotation

$M^{n \times p}$ $N^{n \times k}$

$\hat{A} = \underset{A}{\arg\min} \|M - NA^T\|^2$ subject to $A^TA = I_{k \times k}$

Solution:

$A = \sqrt{tr(AA^T)}$

$\|M - NA^T\|^2$

$= tr(M^TM) - 2tr(M^TNA^T) + tr(AN^TNA^T)$

$= tr(M^TM) - 2tr(M^TNA^T) + tr(N^TNA^TA)$ $(A^TA = I)$

$= tr(M^TM) - 2tr(M^TNA^T) + tr(N^TN)$

Apply SVD: $M^TN = U\Sigma V^T$

$M^TNA = U\Sigma V^T$

$tr(M^TNA^T) = tr(U\Sigma V^TA^T)$. let $A^* = AV$

$\Rightarrow$ $= tr(U\Sigma(AV)^T)$

$= tr(U\Sigma A^{*T})$

$= tr(A^{*T}U\Sigma)$

max this

$A^* = U$

$\hat{A} = UV^T$