SVD AND PCA

Dinh Huu Nguyen, 05/2018

Abstract: a exposition of SVD and PCA, with regard to dimension reduction.

Contents

These notes relate to [1].

1. SVD

Consider a matrix

$$X = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix}$$

of rank l. Assume $l \leq m \leq n$ without loss of generality. Again view X as a linear map with respect to the standard basis $\{e_1, \dots, e_i, \dots, e_m\}$ for $\mathbb{R}^m$ and the standard basis $\{e_1, \dots, e_j, \dots, e_n\}$ for $\mathbb{R}^n$

$$\mathbb{R}^m \xrightarrow{X} \mathbb{R}^n$$
$$u \mapsto Xu$$

Choose an orthogonal collection $\{v_1, \dots v_l\}$ so that $\langle v_1, \dots, v_l\rangle = \operatorname{im}(X)$ and extend it to an orthogonal basis $\{v_1, \dots v_l, v_{l+1}, \dots, v_n\}$ for $\mathbb{R}^n$. Suppose there exist an orthonormal basis $\{u_1, \dots u_l, u_{l+1}, \dots u_m\}$ for $\mathbb{R}^m$ such that $Xu_i = v_i$ for $1 \leq i \leq l$ and $Xu_i = 0$ for $l+1 \leq i \leq m$, that is $\ker(X) = \langle u_{l+1}, \dots, u_m\rangle$. If we define

$$\mathbb{R}^m \xrightarrow{f^t} \mathbb{R}^m$$
$$u_i \mapsto e_i$$
$$\alpha_1 u_1 + \dots + \alpha_m u_m \mapsto \alpha_1 e_1 + \dots + \alpha_m e_m$$

and

$$\mathbb{R}^m \xrightarrow{g} \mathbb{R}^n$$
$$e_i \mapsto \begin{cases} ||v_i||e_i & \text{for } 1 \leq i \leq l \\ 0 & \text{for } l+1 \leq i \leq m \end{cases}$$

and

$$\mathbb{R}^n \xrightarrow{h} \mathbb{R}^n$$
$$e_j \mapsto \frac{v_j}{||v_j||}$$
$$\beta_1 e_1 + \cdots + \beta_n e_n \mapsto \beta_1 \frac{v_1}{||v_1||} + \cdots + \beta_n \frac{v_n}{||v_n||}$$

then

$$\begin{aligned} hgf^t(u_i) &= hg(e_i) \\ &= h(||v_i||e_i) \\ &= v_i \\ &= Xu_i \text{ for } 1 \leq i \leq l \end{aligned}$$

$$\begin{aligned} hgf^t(u_i) &= hg(e_i) \\ &= h(0) \\ &= 0 \\ &= Xu_i \text{ for } l+1 \leq i \leq m \end{aligned}$$

Hence $hgf^t = X$

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, "X"] \arrow[d, "f^t"'] & \mathbb{R}^n \\ \mathbb{R}^m \arrow[r, "g"] & \mathbb{R}^n \arrow[u, "h"'] \end{tikzcd}$$

Surely f^t has matrix form

$$U^t = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix}$$

while g has matrix form

$$\Sigma = \begin{pmatrix} ||v_1|| & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & ||v_m|| & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}$$

and h has matrix form

$$V = \begin{pmatrix} \frac{v_{11}}{||v_1||} & \dots & \frac{v_{n1}}{||v_n||} \\ \vdots & \ddots & \vdots \\ \frac{v_{1n}}{||v_1||} & \dots & \frac{v_{nn}}{||v_n||} \end{pmatrix}$$

$$\begin{tikzcd} \mathbb{R}^m \arrow[r, "X"] \arrow[d, "U^t"'] & \mathbb{R}^n \\ \mathbb{R}^m \arrow[r, "\Sigma"] & \mathbb{R}^n \arrow[u, "V"'] \end{tikzcd}$$

This decomposition $X_{n\times m} = V_{n\times n}\Sigma_{n\times m}U^t_{m\times m}$ is called the singular value decomposition of X. Surely $X^t_{m\times n} = U_{m\times m}\Sigma^t_{m\times n}V^t_{n\times n}$ is the singular value decomposition of X^t.

Example 1.1. (SVD same as EVD) If $\mathbb{R}^m \xrightarrow{X} \mathbb{R}^m$ and $Xu_i = \sigma_i u_i$ for some orthonormal basis $\{u_1, \dots, u_m\}$ for $\mathbb{R}^m$ then σ_i are the eigenvalues and u_i are the eigenvectors of X. It follows from above that

$$\begin{aligned} X &= U\Sigma U^t \\ &= \begin{pmatrix} u_{11} & \dots & u_{m1} \\ \vdots & \ddots & \vdots \\ u_{1m} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_m \end{pmatrix} \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \end{aligned}$$

is the SVD as well as the EVD for X.

2. PCA

If $X_{n\times m}$ has SVD $X_{n\times m} = V_{n\times n}\Sigma_{n\times m}U^t_{m\times m}$ then the Gram matrices XX^t and X^tX both satisfy the conditional statement in example 1.1 and have SVDs (same as EVDs)

$$X^t_{m\times n}X_{n\times m} = U_{m\times m}\Sigma^t_{m\times n}\Sigma_{n\times m}U^t_{m\times m}$$

$$X_{n\times m}X^t_{m\times n} = V_{n\times n}\Sigma_{n\times m}\Sigma^t_{m\times n}V^t_{n\times n}$$

We can rewrite x_j with respect to the orthonormal basis $\{u_1, \dots, u_m\}$ as

$$\begin{aligned} x'_j &= (\langle u_1, x_j\rangle, \dots \langle u_m, x_j\rangle) \\ &= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} x_j \\ &= U^t x_j \end{aligned}$$

and

$$\begin{aligned} U^t_{m\times m}X^t_{m\times n} &= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \dots & x_{nm} \end{pmatrix} \\ &= \begin{pmatrix} x'_{11} & \dots & x'_{n1} \\ \vdots & \ddots & \vdots \\ x'_{1m} & \dots & x'_{nm} \end{pmatrix} \\ &= X'^t_{m\times n} \end{aligned}$$

so that the $X'_1, \dots, X'_m$ have maximal variances $\sigma_1^2, \dots, \sigma_m^2$. Their corresponding singular vectors u_i are call *principal components*.

3. Dimensionality Reduction Without Loss

If the $\sigma_1, \dots, \sigma_l$ are nonzero and the $\sigma_{l+1}, \dots, \sigma_n$ are zero then $\mathrm{im}(X^t) = \langle x_1, \dots, x_n\rangle = \langle u_1, \dots, u_l\rangle$ and $\mathrm{im}(X^t)^\perp = \langle x_1, \dots, x_n\rangle^\perp = \langle u_1, \dots, u_l\rangle^\perp = \langle u_{l+1}, \dots, u_m\rangle$. So

$$\begin{aligned} x'_j &= (\langle u_1, x_j\rangle, \dots, \langle u_l, x_j\rangle, \langle u_{l+1}, x_j\rangle, \dots, \langle u_m, x_j\rangle) \\ &= (\langle u_1, x_j\rangle, \dots, \langle u_l, x_j\rangle, 0, \dots, 0) \\ &= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{l1} & \dots & u_{lm} \\ u_{l+1,1} & \dots & u_{l+1,m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} x_j \\ &= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{l1} & \dots & u_{lm} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} x_j \\ &= \begin{pmatrix} U^t_{l\times m} \\ 0_{(m-l)\times m} \end{pmatrix} x_j \end{aligned}$$

where

$$U^t_{l\times m} = \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{l1} & \dots & u_{lm} \end{pmatrix}$$

Doing this for all x_j can be written as

$$
\begin{aligned}
U^t_{m\times m}X^t_{m\times n} &= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{l1} & \dots & u_{lm} \\ u_{l+1,1} & \dots & u_{l+1,m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \dots & u_{mm} \end{pmatrix} \begin{pmatrix} x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \dots & x_{nm} \end{pmatrix} \\
&= \begin{pmatrix} u_{11} & \dots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{l1} & \dots & u_{lm} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1m} & \dots & x_{nm} \end{pmatrix} \\
&= \begin{pmatrix} U^t_{l\times m} \\ 0_{(m-l)\times m} \end{pmatrix} X^t \\
&= \begin{pmatrix} x'_{11} & \dots & x'_{n1} \\ \vdots & \ddots & \vdots \\ x'_{1l} & \dots & x'_{nl} \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \\
&= X'^t_{m\times n}
\end{aligned}
$$

or as $X_{n\times m} \begin{pmatrix} U_{m\times l} & 0_{m\times(m-l)} \end{pmatrix} = X'_{n\times m}$ in data science.

If one keeps only the l nonzero coordinates of x'_j then

$$x'_j = U^t_{l\times m} x_j$$

Doing this for all x_j can be written as $U^t_{l\times m}X^t_{m\times n} = X'^t_{l\times n}$, or as $X_{n\times m}U_{m\times l} = X'_{n\times l}$ in data science.

$$
\begin{tikzcd}
\mathbb{R}^m \arrow[d, "U^t_{m\times m}"'] \arrow[dr, "U^t_{l\times m}"] & \mathbb{R}^n \arrow[l, "X^t_{m\times n}"'] \arrow[d, "X'^t_{l\times n}"] \\
\mathbb{R}^m \arrow[r, "\pi_l"] & \mathbb{R}^l
\end{tikzcd}
$$

This dropping of the zero coordinates for all x_i incurs no "loss of information" about X, in the sense that $U_{m\times l}X'^t_{l\times m} = U_{m\times l}U^t_{l\times m}X^t_{m\times n} = X^t_{m\times n}$ that we saw in [1, 2.1].

4. Further Dimensionality Reduction With Loss

In data science, one often arranges the σ_i in decreasing order, chooses a $k < l$ and keeps only the first k nonzero coordinates of x'_i

$$x'_i = V^t_{k\times n} x_i$$

where

$$V^t_{k\times n} = \begin{pmatrix} v_{11} & \dots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{k1} & \dots & v_{kn} \end{pmatrix}$$

Doing this for all x_i can be written as $V^t_{k\times n} X^t_{n\times m} = X'^t_{k\times m}$, or as $X_{m\times n} V_{n\times k} = X'_{m\times k}$ in data science.

$$\begin{tikzcd}
\mathbb{R}^m \arrow[r, "X^t"] \arrow[d, "V^t_{k\times n}X^t = X'^t"'] & \mathbb{R}^n \arrow[d, "V^t"] \arrow[dl, "V^t_{k\times n}"'] \\
\mathbb{R}^k & \mathbb{R}^n \arrow[l, "\pi_k"']
\end{tikzcd}$$

This dropping of the last $n-k$ coordinates for all x_j incurs "loss of information" about X, in the sense that $V_{n\times k} X'^t_{k\times m} = V_{n\times k} V^t_{k\times n} X^t_{n\times m} \neq X^t_{n\times m}$ that we saw in section [1, 3.1]. Again this "loss of information" is quantified by $||V_{n\times k} V^t_{k\times n} X^t_{n\times m} - X^t_{n\times m}||$.

If $\{v_1, \dots, v_n\}$ is not orthonormal?

The fixed precision problem on the other hand is: given $\epsilon > 0$, find $\mathbb{R}^m \xrightarrow{B} \mathbb{R}^n$ of smallest rank k such that $||B - X^t|| < \epsilon$.

Example 4.1. We can rewrite

X	X_1	X_2	X_3
x_1	$\sqrt{2}$	$\sqrt{2}$	0
x_2	$2\sqrt{2}$	$2\sqrt{2}$	0
x_3	0	$\sqrt{2}$	$\sqrt{2}$

with respect to orthonormal basis $\{(1,0,0),(0,1,0),(0,0,1)\}$ as

	X'_1	X'_2	X'_3
x_1	0	1	0
x_2	0	2	0
x_3	0	0	1

with respect to general basis $\{(1,0,0),(\sqrt{2},\sqrt{2},0),(0,\sqrt{2},\sqrt{2})\}$.

References

[1] D. Nguyen, M. Wojnowicz, L. Li, and X. Zhao, *Low Rank Decomposition and Dimension Reduction.*