MATH 271.1: Statistical Methods

Supplementary Notes on Principal Components Analysis

Jakov Ivan S. Dumbrique

2 The Singular Value Decomposition

Consider a matrix $\mathbf{A} \in \mathbb{R}^{n \times p}$. Singular Value Decomposition (SVD) (Stewart, 1993) is one of the most well-known and effective matrix decomposition methods. We want to diagonalize this $\mathbf{A}$, but not by $\mathbf{P^{-1}AP}$. The eigenvectors in $\mathbf{P}$ have three big problems: They are usually not orthogonal, there are not always enough eigenvectors, and $\mathbf{Ax} = \lambda\mathbf{x}$ requires $\mathbf{A}$ to be a square matrix. In SVD, we will be able to find the right bases for the four fundamental subspaces.

The SVD has two different forms, i.e. complete and incomplete.

The "complete SVD" decomposes the matrix as:

$$\mathbf{A} = \mathbf{U\Sigma V^T},$$

where

$\mathbf{U} \in \mathbb{R}^{n \times n}$ is orthonormal
$\mathbf{\Sigma} \in \mathbb{R}^{n \times p}$ is diagonal
$\mathbf{V} \in \mathbb{R}^{p \times p}$ is orthonormal

The columns of $\mathbf{U}$ are called the *left singular vectors* of $\mathbf{A}$.
The columns of $\mathbf{V}$ are called the *right singular vectors* of $\mathbf{A}$.
In complete SVD, the Σ is a rectangular diagonal matrix. In cases $n > p$ and $n < p$, this matrix is in the forms:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_p \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}, \text{ and } \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_n & 0 & \dots & 0 \end{bmatrix},$$

The diagonal entries of $\mathbf{\Sigma}$, denoted by σ_i are nonnegative and are called the *singular values* of $\mathbf{A}$. They are ordered such that the largest singular value, σ_1 is placed in the $(1, 1)$ entry of $\mathbf{\Sigma}$, and the other singular values are placed down the diagonal, and satisfy $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_l \geq 0$ such that $l = \min(n, p)$.

The "incomplete SVD" decomposes the matrix as:

$$\mathbf{A} = \mathbf{U\Sigma V^T},$$

where

$\mathbf{U} \in \mathbb{R}^{n \times l}$ is orthonormal
$\mathbf{\Sigma} \in \mathbb{R}^{l \times l}$ is diagonal
$\mathbf{V} \in \mathbb{R}^{p \times l}$ is orthonormal,

such that $l := \min(n, p)$. The columns of $\mathbf{U}$ and the columns of $\mathbf{V}$ are still called the *left singular vectors* and the *right singular vectors* of $\mathbf{A}$, respectively. In incomplete SVD, the Σ is a square diagonal matrix whose main diagonal includes the *singular values* of $\mathbf{A}$. The matrix Σ is in the form:

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_l \end{bmatrix}.$$

Note that in both complete and incomplete SVD, the left singular vectors are orthonormal and the right singular vectors are also orthonormal; therefore, $\mathbf{U}$ and $\mathbf{V}$ are both orthogonal matrices so:

$$\mathbf{U}^\mathbf{T}\mathbf{U} = \mathbf{I},$$
$$\mathbf{V}^\mathbf{T}\mathbf{V} = \mathbf{I}.$$

If these orthogonal matrices are not truncated and thus are square matrices, e.g., for complete SVD, we also have:

$$\mathbf{U}\mathbf{U}^\mathbf{T} = \mathbf{I},$$
$$\mathbf{V}\mathbf{V}^\mathbf{T} = \mathbf{I}.$$

If $\mathbf{A}$ has r non-zero singular values, counted according to their multiplicity (i.e. if $\mathbf{\Sigma}$ has r non-zero diagonal entries), then $\mathbf{A}$ has rank r. That is, the rank of $\mathbf{A}$ is equal to the number of non-zero singular values, when their multiplicities are accounted for. Also, for $j > r$, the singular value σ_j has a value of 0.

Theorem 1. In both complete and incomplete SVD, the singular values of an $n \times p$ matrix $\mathbf{A}$ are the square roots of the eigenvalues of $\mathbf{A}^\mathbf{T}\mathbf{A}$. The right singular vectors are the eigenvectors of $\mathbf{A}^\mathbf{T}\mathbf{A}$.

<u>Proof</u>: Suppose that the $n \times p$ matrix $\mathbf{A}$ has a singular value decomposition as follows:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}.$$

Therefore,

$$\begin{aligned}
\mathbf{A}^\mathbf{T}\mathbf{A} &= \left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}\right)^\mathbf{T}\left(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T}\right) \\
&= \mathbf{V}\mathbf{\Sigma}^\mathbf{T}\mathbf{U}^\mathbf{T}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathbf{T} \\
&= \mathbf{V}(\mathbf{\Sigma}^\mathbf{T}\mathbf{\Sigma})\mathbf{V}^\mathbf{T} \\
&= \mathbf{V}(\mathbf{\Sigma}^\mathbf{T}\mathbf{\Sigma})\mathbf{V}^{-\mathbf{1}}
\end{aligned}$$

Thus, $\mathbf{A}^\mathbf{T}\mathbf{A}$ is similar to $\mathbf{\Sigma}^\mathbf{T}\mathbf{\Sigma}$, which implies that they have the same eigenvalues. Since $\mathbf{\Sigma}^\mathbf{T}\mathbf{\Sigma}$ is a square diagonal matrix, its eigenvalues are in fact the diagonal entries, which are the squares of the singular values.

We have effectively performed an eigenvalue decomposition for $\mathbf{A}^\mathbf{T}\mathbf{A}$. Indeed, since $\mathbf{A}^\mathbf{T}\mathbf{A}$ is symmetric, this is an orthogonal diagonalization, and thus the eigenvectors of

$\mathbf{A^T A}$ are the columns of $\mathbf{V}$, which are the right singular vectors of $\mathbf{A}$.

Theorem 2. Given an $n \times p$ matrix $\mathbf{A}$ of rank r, then the left and right singular vectors in its complete singular value decomposition

$$\mathbf{A} = \mathbf{U \Sigma V^T}$$

form orthonormal bases for the four fundamental subspaces. In particular,

$$\begin{aligned}
\mathbf{u_1}, \ldots, \mathbf{u_r} &\text{ is an orthonormal basis for the } \textbf{column space } Col(\mathbf{A}) \\
\mathbf{u_{r+1}}, \ldots, \mathbf{u_n} &\text{ is an orthonormal basis for the } \textbf{left nullspace } N(\mathbf{A^T}) \\
\mathbf{v_1}, \ldots, \mathbf{v_r} &\text{ is an orthonormal basis for the } \textbf{row space } Col(\mathbf{A^T}) \\
\mathbf{v_{r+1}}, \ldots, \mathbf{v_p} &\text{ is an orthonormal basis for the } \textbf{nullspace } N(\mathbf{A})
\end{aligned}$$

<u>Proof</u>:

1. $\mathbf{u_1}, \ldots, \mathbf{u_r}$ is an orthonormal basis for the **column space** $Col(\mathbf{A})$
   for $1 \leq i \leq r$, we have

$$\begin{aligned}
\mathbf{Av_i} &= \mathbf{U \Sigma V^T v_i} \\
&= \mathbf{U \Sigma e_i} \\
&= \mathbf{U} \sigma_i \mathbf{e_i} \\
&= \sigma_i \mathbf{u_i}
\end{aligned}$$

   Thus, the vectors $\mathbf{u_i}$ are in $Col(\mathbf{A})$, and because they are orthonormal, then the vectors $\mathbf{u_i}$ are also linearly independent. Since $dim\ Col(\mathbf{A}) = r$, then the linearly independent set $\{\mathbf{u_1}, \ldots, \mathbf{u_r}\}$ must form an orthonormal basis for $Col(\mathbf{A})$.

2. $\mathbf{v_{r+1}}, \ldots, \mathbf{v_p}$ is an orthonormal basis for the **nullspace** $N(\mathbf{A})$
   for $j > r$, we have

$$\begin{aligned}
\mathbf{Av_j} &= \mathbf{U \Sigma V^T v_j} \\
&= \mathbf{U \Sigma e_j} \\
&= \mathbf{U} \sigma_j \mathbf{e_j} \\
&= \mathbf{U} 0 \mathbf{e_j} \\
&= \mathbf{0}
\end{aligned}$$

   Thus, the vectors $\mathbf{v_j}$ are in $N(\mathbf{A})$, and because they are orthonormal, then the vectors $\mathbf{v_j}$ are also linearly independent. Since $dim\ N(\mathbf{A}) = p - r$, then the linearly independent set $\{\mathbf{v_{r+1}}, \ldots, \mathbf{v_p}\}$ must form an orthonormal basis for $N(\mathbf{A})$.

To find bases for $N(\mathbf{A^T})$ and $Col(\mathbf{A^T})$, just notice that $\mathbf{A^T} = (\mathbf{U \Sigma V^T})^\mathbf{T} = \mathbf{V \Sigma^T U^T}$. This is a SVD for $\mathbf{A^T}$, because the diagonal entries of $\mathbf{\Sigma}$ and $\mathbf{\Sigma^T}$ are the same, and $\mathbf{V}$ and $\mathbf{U}$ are orthogonal. Applying the same reasoning as above to $\mathbf{A^T} = \mathbf{V \Sigma^T U^T}$, we find that the first r columns of $\mathbf{V}$ are a basis for $Col(\mathbf{A^T})$ and the last $n - r$ columns of $\mathbf{U}$ are a basis for $N(\mathbf{A^T})$.