1 Singular Value Decomposition (SVD)

Diagonalization(PDP^{-1}) or orthogonal diagonalization(PDP^T) are useful when you want to understand what kind of properties your matrix A has. However, not all matrices are diagonalizable. Here is a very simple non-example: $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Moreover, we would like to also talk something about non-square matrices. In order to handle these cases, we have a decomposition of A such that it is good enough and you always have this decomposition whatever your matrix A is.

1.1 Singular Values

Recall that a symmetric matrix is always orthogonally diagonalizable and it's good. In fact, there is a natural way to make a symmetric matrix out of any $m \times n$ (note that it does not need to be square) matrix A. That is, A^TA. In some sense, A^TA seems to be a good way to think about "A squared" (note that this is not even defined unless $m = n$) if you recall that $\mathbf{x}^T A^T A\mathbf{x} = ||A\mathbf{x}||^2$. So, we can presume that the square roots of the eigenvalues of A^TA would be good candidates for "eigenvalues" of A. Note that there cannot be any eigen-things for a non-square matrix A and we will use another name, **singular values** of A.

> The eigenvalues of A^TA is nonnegative, say $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ are eigenvalues. We take the (positive) square root of each eigenvalue and call it as a singular value, denoted by $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

Now, a miracle happens.

1.2 The Singular Value Decomposition

1.2.1 Theorem 9

As A^TA is an $n \times n$ symmetric matrix, according to the Spectral Theorem, we have an orthonormal basis consisting of eigenvectors. Let $v_1, \cdots, v_n$ be the vectors such that the corresponding eigenvalues satisfy $\lambda_1 \geq \cdots \geq \lambda_n \geq 0$. If there are r nonzero numbers, then $\{Av_1, \cdots, Av_r\}$ is an ortho**gonal** basis for Col A and rk$A = r$.

1.2.2 Theorem 10

Let A be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix $\Sigma = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ where D is an $r \times r$ diagonal matrix with nonnegative entries, and there exist an $m \times m$ orthogonal matrix U and an $n \times n$ orthogonal matrix V such that

$$A = U\Sigma V^T.$$

2 Why is Linear Algebra (sometimes) useful?

2.1 Statistics

Some data of N people can be expressed as following: $\mathbf{X} = \begin{bmatrix} w_1 & w_2 & \cdots & w_N \\ h_1 & h_2 & \cdots & h_N \\ d_1 & d_2 & \cdots & d_N \\ b_1 & b_2 & \cdots & b_N \end{bmatrix}$. The sample mean of $\mathbf{X}$ is $\mathbf{X} \begin{bmatrix} 1/N \\ 1/N \\ \vdots \\ 1/N \end{bmatrix}$.

Mean-deviation form of $\mathbf{X}$, denoted by $\mathbf{B}$, is the matrix with new column obtained by subtracting the sample mean from each $\mathbf{X}$'s column. The sample covariance matrix is

$$S = \frac{1}{N-1}\mathbf{B}\mathbf{B}^T.$$

S's diagonal entries are called variances and non-diagonal entries are called covariances. In statistics, one uses the covariance to measure how closely two variables are related. Finally, we define Total variance to be tr$S$.

2.2 Data Analysis

We will take a look at a tiny example among (tons of) applications of SVD in data analysis: Principal Component Analysis (PCA). Suppose you want to transform your data linearly via an orthogonal matrix; $\mathbf{X} = P\mathbf{Y}$ or $\mathbf{Y} = P^T\mathbf{X}$. Then, the covariance matrix for $\mathbf{Y}$ becomes P^TSP. We will make this be diagonal and then do **principal component analysis**.