The Singular Value Decomposition

The Singular Values of a Matrix

Let A be any real $m \times n$ matrix. Then $\mathrm{A}^T\mathrm{A}$ is a symmetric $n \times n$ matrix since

$$\left(\mathrm{A}^T\mathrm{A}\right)^T = \mathrm{A}^T\mathrm{A}^{TT} = \mathrm{A}^T\mathrm{A}$$

There is a theorem which says that all square matrices that are symmetric have an orthonormal set of eigenvectors that span $\mathbb{R}^n$ (in fact only square matrices that are symmetric have this property, but that's beside the point). Let $\{\mathrm{v}_1, \ldots, \mathrm{v}_n\}$ be an orthonormal set of eigenvectors of $\mathrm{A}^T\mathrm{A}$ and let the corresponding eigenvalues be $\sigma_1^2, \ldots, \sigma_n^2$. Then

$$|\mathrm{A}\mathrm{v}_i|^2 = (\mathrm{A}\mathrm{v}_i)^T\mathrm{A}\mathrm{v}_i = \mathrm{v}_i^T\mathrm{A}^T\mathrm{A}\mathrm{v}_i = \mathrm{v}_i^T\sigma_i^2\mathrm{v}_i = \sigma_i^2|\mathrm{v}_i|^2 = \sigma_i^2$$

$$|\mathrm{A}\mathrm{v}_i| = \sigma_i$$

Definition: The singular values of an $m \times n$ matrix A are the square roots σ_i of the eigenvalues σ_i^2 of $\mathrm{A}^T\mathrm{A}$, which is also the length of the vectors $\mathrm{A}\mathrm{v}_i$ where v_i are the eigenvectors of $\mathrm{A}^T\mathrm{A}$.

The singular values are arranged in descending order, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$.

The Singular Value Decomposition

Let A be any real $m \times n$ matrix with rank r (the rank being the maximum number of independent columns of A OR the maximum number of independet rows of A). A can be written on the form

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

where Σ is an $m \times n$ matrix on the form

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \end{bmatrix}$$

V is an $n \times n$ orthogonal matrix with the orthonormal set of eigenvectors of $\mathrm{A}^T\mathrm{A}$ as columns

$$\mathrm{V} = [\mathrm{v}_1, \, \ldots, \mathrm{v}_n]$$

and U is an $m \times m$ orthogonal matrix with an orthonormal basis $\{\mathrm{u}_1, \, \ldots, \mathrm{u}_m\}$ of $\mathbb{R}^m$ as columns

$$\mathrm{U} = [\mathrm{u}_1, \, \ldots, \mathrm{u}_m]$$

where the first r columns are given by

$$\mathrm{u}_i = \frac{1}{\sigma_i}\mathrm{A}\mathrm{v}_i. \;\; i = 1, \, \ldots, r$$

while the last $m - r$ columns are arbitrary (although it is convenient to choose the last columns such that the full set is an orthonormal basis of $\mathbb{R}^m$ so that U is unitary).

The Reduced Singular Value Decomposition and the Pseudoinverse

We can write U and V as

$$\mathrm{U} = \begin{bmatrix} \mathrm{U}_r & \mathrm{U}_{m-r} \end{bmatrix}$$

$$\mathrm{V} = \begin{bmatrix} \mathrm{V}_r & \mathrm{V}_{n-r} \end{bmatrix}$$

where the columns of e.g U_r are the first r columns of U while the columns of U_{m-r} are the last $m - r$ columns of U. Then. because the last $m - r$ rows and the last $n - r$ columns of Σ contain only zeros, we have

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

$$= [\mathrm{U}_r\ \mathrm{U}_{m-r}]\Sigma\begin{bmatrix}\mathrm{V}_r^T \\ \mathrm{V}_{n-r}^T\end{bmatrix} = \mathrm{U}_r\mathrm{D}\mathrm{V}_r^T$$

where D is an $r \times r$ diagonal matrix containing the singular values of A.

$$\mathrm{D} = \begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$$

$\mathrm{A} = \mathrm{U}_r\mathrm{D}\mathrm{V}_r^T$ is called the reduced singular value decomposition of A. Note that even though U_r and V_r have orthogonal columns they are not orthogonal matrices since they are not square; they have shape $m \times r$ and $n \times r$ respectively. If they were orthogonal matrices then the inverse of A would be $\mathrm{V}_r\mathrm{D}^{-1}\mathrm{U}_r^T$. We define the pseudoinverse A^+ of A to be

$$\mathrm{A}^+ \equiv \mathrm{V}_r\mathrm{D}^{-1}\mathrm{U}_r^T$$

Application to Ordinary Least Squares regression

The OLE solution can be written as

$$\tilde{\mathrm{y}} = \mathrm{X}\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

Suppose we have an SVD of $\mathrm{X} = \mathrm{U}\Sigma\mathrm{V}^T$.

$$\mathrm{X}^T\mathrm{X} = \left(\mathrm{U}\Sigma\mathrm{V}^T\right)^T\mathrm{U}\Sigma\mathrm{V}^T$$

$$= \mathrm{V}\Sigma^T\mathrm{U}^T\mathrm{U}\Sigma\mathrm{V}^T = \mathrm{V}\Sigma^T\Sigma\mathrm{V}^T$$

Inserting this into the expression for the OLE solution

$$\tilde{\mathrm{y}} = \mathrm{X}\left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{y}$$

$$= \mathrm{U}\Sigma\mathrm{V}^T\left(\mathrm{V}\Sigma^T\Sigma\mathrm{V}^T\right)^{-1}\left(\mathrm{U}\Sigma\mathrm{V}^T\right)^{\mathrm{T}}\mathrm{y} = \mathrm{U}\Sigma\mathrm{V}^T\left(\mathrm{V}\Sigma^T\Sigma\mathrm{V}^T\right)^{-1}\mathrm{V}\Sigma^{\mathrm{T}}\mathrm{U}^{\mathrm{T}}\mathrm{y}$$

$$= \mathrm{U}\Sigma\mathrm{V}^T\left(\mathrm{V}^T\right)^{-1}\Sigma^{-1}\left(\Sigma^{\mathrm{T}}\right)^{-1}\mathrm{V}^{-1}\mathrm{V}\Sigma^{\mathrm{T}}\mathrm{U}^{\mathrm{T}}\mathrm{y} = \mathrm{U}\Sigma\Sigma^{-1}\left(\Sigma^{\mathrm{T}}\right)^{-1}\Sigma^{\mathrm{T}}\mathrm{U}^{\mathrm{T}}\mathrm{y}$$

$$= \mathrm{UU^T y}$$

(mfw i just realized Σ is not invertible ._.)