MA 511, Session 42

Singular Value Decomposition

Let A be a $m \times n$ real matrix of rank r.

Theorem: There exist an orthogonal $m \times m$ matrix Q_1, an orthogonal $n \times n$ matrix Q_2, and a $m \times n$ matrix $\Sigma = (\sigma_{ij})$ with all its coefficients equal to zero, except for $\sigma_{11}, \ldots, \sigma_{rr}$ that are positive real numbers, such that $A = Q_1 \Sigma Q_2^T$. Moreover, the columns of Q_1 are an orthonormal basis of $\mathbb{R}^m$ of eigenvectors of AA^T, and those of Q_2 are an orthonormal basis of $\mathbb{R}^n$ of eigenvectors of $A^T A$.

Definition: The singular values of A, $\sigma_1, \ldots, \sigma_r$, are the square roots of the nonzero eigenvalues of AA^T (which are the same as those of $A^T A$), $\sigma_j = \sqrt{\sigma_{jj}}$, $1 \le j \le r$.

Remark: If A is positive definite, the singular value decomposition is just the unitary diagonalization $A = Q \Lambda Q^T$ we have already seen.

Proof: (of theorem) Without loss of generality, assume $m \geq n$. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $A^T A$ and let $x_1, \ldots, x_n$ be an orthonormal set of eigenvectors of $A^T A$; place these as columns of Q_2. Then,

$$A^T A x_j = \lambda_j x_j, \text{ and } \lambda_j \geq 0$$

since

$$(*) \qquad \lambda_j = \lambda_j x_j^T x_j = x_j^T A^T A x_j = \|A x_j\|^2.$$

Let now $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ and let $\sigma_j = \sqrt{\lambda_j}$. Then, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ and $\sigma_j = 0$ for $r + 1 \leq j \leq n$ since $\operatorname{rank}(A^T A) = r$ as well.

For $1 \leq j \leq r$, define now $q_j = \frac{A x_j}{\sigma_j}$. By $(*)$, these vectors have unit length; moreover, for $i \neq j$,

$$q_i^T q_j = \frac{x_i^T A^T A x_j}{\sigma_i \sigma_j} = \frac{\lambda_j}{\sigma_i \sigma_j} x_i^T x_j = 0,$$

so that they are an orthonormal set. Extend this set to an orthonormal basis of $\mathbb{R}^n$ (using Gram-Schmidt if necessary), and take this basis as columns of Q_1. Then, the coefficient in the i-th row and the j-th column of $Q_1^T A Q_2$ is

$$q_i^T A x_j = \begin{cases} 0, & j > r \quad (\text{since then } Ax_j = 0), \\ q_i^T \sigma_j q_j = \sigma_j \delta_{ij}, & j \le r, \end{cases}$$

where δ_{ij} is Kronecker's symbol. Hence $Q_1^T A Q_2 = \Sigma$, and $A = Q_1 \Sigma Q_2^T$. All that is left to conclude the proof is to check that the columns of Q_1 are, in fact, eigenvectors of AA^T:

$$AA^T = Q_1 \Sigma Q_2^T Q_2 \Sigma^T Q_1^T = Q_1 (\Sigma \Sigma^T) Q_1^T,$$

where $\Sigma\Sigma^T$ is a $m \times m$ diagonal matrix. It follows that such diagonal matrix contains the eigenvalues of AA^T on its diagonal, and Q_1 has corresponding eigenvectors for columns, thus concluding the proof.

Example: Application to Image Processing.

Suppose that A is a 1000×1000 matrix containing the data for an image of 1000×1000 pixels (i.e. one million data values). Let $A = Q_1 \Sigma Q_2^T$ be the singular value decomposition of A. Suppose we replace Σ by a matrix $\tilde{\Sigma}$ which keeps only the largest singular values of A, say the first 60, for example. Thus, $\tilde{\sigma}_{jj} = \sigma_{jj}$ for $1 \le j \le 60$ and $\tilde{\sigma}_{jj} = 0$ for $61 \le j \le 1000$. Let $u_1, \ldots, u_{1000}$ and $v_1, \ldots, v_{1000}$ denote, respectively, the columns of Q_1 and Q_2.

We have

$$A \approx Q_1 \tilde{\Sigma} Q_2^T = \sigma_1 u_1 v_1^T + \cdots + \sigma_{60} u_{60} v_{60}^T$$

requires now just $60 + 60 \times 2000 = 120{,}060$ data values, with savings of nearly 78%.

Example: Find a singular value decomposition of $A = (\,1 \quad 1 \quad 1\,)$.

<u>Solution</u>: $AA^T = 3$ and $A^TA = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$. It follows that $Q_1 = 1$. The characteristic polynomial of A^TA is $p(\lambda) = -\lambda^2(\lambda - 3)$, so that the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = \lambda_3 = 0$. We find eigenvectors for $\lambda = 3$ by solving $(A^TA - 3\mathbb{I})x = 0$ and find that $S_3 = \text{ span}\,\{(1,1,1)^T\}$. Similarly, we find eigenvectors for $\lambda = 0$ by solving $A^TAx = 0$ and find that $S_0 = \text{ span}\,\{(1,1,-2)^T, (1,-2,1)^T\}$, where we chose an orthogonal basis for S_0 (otherwise, we would use

Gram-Schmidt on *any* basis of S_0 to produce an orthonormal one). We normalize these eigenvectors and place them as columns of Q_2.

Thus,

$$Q_2 = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix}.$$

Finally,

$$\begin{aligned} \Sigma &= Q_1^T A Q_2 \\ &= 1 \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{3} & 0 & 0 \end{pmatrix}, \end{aligned}$$

where $\sigma_1 = \sqrt{\lambda_1} = \sqrt{3}$ is the only singular value of A.