Principal component analysis

Ingo Blechschmidt
December 17th, 2014

December 17th, 2014

Outline

1. Theory
   - Singular value decomposition
   - Pseudoinverses
   - Low-rank approximation

Singular value decomposition

Let $A \in \mathbb{R}^{n\times m}$. Then there exist

- numbers $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$,
- an orthonormal basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$ of $\mathbb{R}^m$, and
- an orthonormal basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of $\mathbb{R}^n$,

such that

$$A\mathbf{v}_i = \sigma_i \mathbf{w}_i, \quad i = 1, \ldots, m.$$

In matrix language:

$$\begin{aligned} & A = W\Sigma V^t, \\ \text{where} \quad & V = (\mathbf{v}_1 | \ldots | \mathbf{v}_m) \in \mathbb{R}^{m\times m} \text{ orthogonal}, \\ & W = (\mathbf{w}_1 | \ldots | \mathbf{w}_n) \in \mathbb{R}^{n\times n} \text{ orthogonal}, \\ & \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m) \in \mathbb{R}^{n\times m}. \end{aligned}$$

- The singular value decomposition (SVD) exists for any real matrix, even rectangular ones.

- The singular values σ_i are unique.

- The basis vectors are not unique.

- If A is orthogonally diagonalizable with eigenvalues λ_i (for instance, if A is symmetric), then $\sigma_i = |\lambda_i|$.

- $\|A\|_{\text{Frobenius}} = \sqrt{\sum_{ij} A_{ij}^2} = \sqrt{\text{tr}(A^t A)} = \sqrt{\sum_i \sigma_i^2}$.

- There exists a generalization to complex matrices. In this case, the matrix A can be decomposed as $W \Sigma V^\star$, where $V^\star$ is the complex conjugate of V^t, and W and V are unitary matrices.

- The singular value decomposition can also be formulated in a basis-free manner as a result about linear maps between finite-dimensional Hilbert spaces.

Existence proof (sketch):

1. Consider the eigenvalue decomposition of the symmetric and positive-semidefinite matrix $A^t A$: We have an orthonormal basis $\mathbf{v}_i$ of eigenvectors corresponding to eigenvalues λ_i.

2. Set $\sigma_i := \sqrt{\lambda_i}$.

3. Set $\mathbf{w}_i := \frac{1}{\sigma_i} A\mathbf{v}_i$ (for those i with $\lambda_i \neq 0$).

4. Then $A\mathbf{v}_i = \sigma_i \mathbf{w}_i$ holds trivially.

5. The $\mathbf{w}_i$ are orthonormal: $(\mathbf{w}_i, \mathbf{w}_j) = \frac{1}{\sigma_i \sigma_j}(A^t A\mathbf{v}_i, \mathbf{v}_j) = \frac{\lambda_i \delta_{ij}}{\sigma_i \sigma_j}$.

6. If necessary, extend the $\mathbf{w}_i$ to an orthonormal basis.

This proof gives rise to an algorithm for calculating the SVD, but unless $A^t A$ is small, it has undesirable numerical properties. Since the 1960ies, there exists a stable iterative algorithm by Golub and van Loan.

The pseudoinverse of a matrix

Let $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the solutions to the optimization problem

$$\|A\mathbf{x} - \mathbf{b}\|_2 \longrightarrow \min$$

under $\mathbf{x} \in \mathbb{R}^m$ are given by

$$\mathbf{x} = A^+\mathbf{b} + V\begin{pmatrix} 0 \\ \star \end{pmatrix},$$

where $A = W\Sigma V^t$ is the SVD and

$$A^+ = V\Sigma^+W^t,$$
$$\Sigma^+ = \operatorname{diag}(\sigma_1^{-1}, \ldots, \sigma_m^{-1}).$$

- In the formula for Σ^+, set $0^{-1} := 0$.
- The pseudoinverse can be used for interpolation: Let data points $(x_i, y_i) \in \mathbb{R}^2$, $1 \leq i \leq N$, be given. Want to find a polynomial $p(z) = \sum_{k=0}^{n} \alpha_i z^i$, $n \ll N$, such that

$$\sum_{i=1}^{N} |p(x_i) - y_i|^2 \longrightarrow \min.$$

In matrix language, this problem is written

$$\|A\mathbf{u} - \mathbf{y}\|_2 \longrightarrow \min$$

where $\mathbf{u} = (\alpha_0, \ldots, \alpha_N)^T \in \mathbb{R}^{n+1}$ and

$$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \cdots & x_N^n \end{pmatrix} \in \mathbb{R}^{N\times(n+1)}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \in \mathbb{R}^N.$$

Low-rank approximation

Let $A = W\Sigma V^t \in \mathbb{R}^{n\times m}$ and $1 \le r \le n, m$. Then a solution to the optimization problem

$$\|A - M\|_{\text{Frobenius}} \longrightarrow \min$$

under all matrices M with $\operatorname{rank} M \le r$ is given by

$$\begin{aligned} M &= W\Sigma_r V^t, \\ \text{where} \quad \Sigma_r &= \operatorname{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0). \end{aligned}$$

The approximation error is

$$\|A - W\Sigma_r V^t\|_{\mathrm{F}} = \sqrt{\sigma_{r+1}^2 + \cdots + \sigma_m^2}.$$