Lecture 2c: Singular Value Decomposition (SVD of) Data and Systems

Lecturer: Jeffrey Varner

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

Topics

- **Singular Value Decomposition (SVD)** is a factorization of a matrix into the product of three matrices: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values of $\mathbf{A}$. The SVD is used to analyze the structure of a matrix, and is used in many applications, including data compression, image processing, and control theory.
- **Principle Component Analysis (PCA)** is a method for reducing the dimensionality of data by projecting it onto a lower-dimensional subspace. The principal components are the eigenvectors of the covariance matrix of the data, and the principal component scores are the projections of the data onto the principal components. We can compute the principal components using the SVD of the covariance matrix.

1 Introduction

In this lecture, we will discuss the singular value decomposition (SVD) of data and systems, and a method for reducing the dimensionality of data called principal component analysis (PCA), which is similar to singular value decomposition. The SVD is a fundamental matrix decomposition that is used in many areas of science and engineering. The SVD is a generalization of the eigenvalue decomposition and is used to analyze the structure of a matrix, for non-square matrices. The SVD is used in a huge variety of unsupervised learning type applications, e.g., understanding gene expression data (1, 2), the structure of chemical reaction networks (3), in process control applications (4), and analyis of various type of networks (5, 6). Similarly, PCA is a widely used method across many fields and applicatrions, e.g., drug discovery (7).

2 Singular Value Decomposition (SVD)

Singular value decomposition (SVD), originally developed in 1870s (8) is a matrix factorization technique that is based on the eigendecomposition of a matrix. Suppose we have a matrix $A \in \mathbb{R}^{m \times n}$. The SVD of $\mathbf{A}$ is a factorization of the form: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ are orthogonal matrices, i.e., $\mathbf{U} \cdot \mathbf{U}^T = \mathbf{I}$ and $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$ is a diagonal matrix containing the singular values $\sigma_i = \Sigma_{ii}$. The matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ can be decomposed as:

$$\mathbf{A} = \sum_{i=1}^{r_{\mathbf{A}}} \sigma_i \cdot (\mathbf{u}_i \otimes \mathbf{v}_i)$$

where $r_{\mathbf{A}}$ is the rank of matrix $\mathbf{A}$ and $\otimes$ denotes the outer product. The outer product $\hat{\mathbf{A}}_i = \mathbf{u}_i \otimes \mathbf{v}_i$ is a rank-1 matrix with elements:

$$\hat{a}_{jk} = u_j v_k \qquad j = 1, 2, \ldots, n \text{ and } k = 1, 2, \ldots, m$$

The vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left (right) singular vectors, and σ_i are the singular values (ordered). Singular value decomposuition is a special sort of eigendecomposition, thus, we could use QR-itereation to compute the SVD. The columns of $\mathbf{U}$ are eigenvectors of $\mathbf{AA}^T$, the columns of $\mathbf{V}$ are eigenvectors of $\mathbf{A}^T\mathbf{A}$ and the singular values σ_i are the square roots of the eigenvalues of the matrix products $\mathbf{AA}^T$ or $\mathbf{A}^T\mathbf{A}$.

3 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a method for reducing the dimensionality of data by projecting it onto a lower-dimensional subspace. The principal components are the eigenvectors of the covariance matrix of the data, and the principal component scores are the projections of the data onto the principal components.

Let's return to the idea of dimensionality reduction of the covariance matrix that we introduced when exploring eigendecomposition. Suppose we have a dataset $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ where each $\mathbf{x}_i \in \mathbb{R}^m$ is an m-dimensional feature vector. The covariance of feature vectors i and j, denoted as $\mathrm{cov}(\mathbf{x}_i, \mathbf{x}_j)$, is an $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ real-valued symmetric matrix $\mathbf{\Sigma} \in \mathbb{R}^{n \times n}$ with elements:

$$\Sigma_{ij} = \mathrm{cov}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_i \, \sigma_j \, \rho_{ij} \qquad \text{for} \quad i, j \in \mathcal{D} \tag{1}$$

where σ_i denote the standard deviation of the feature vector $\mathbf{x}_i$, σ_j denote the standard deviation of the feature vector $\mathbf{x}_j$, and ρ_{ij} denotes the correlation between features i and j in the dataset $\mathcal{D}$. The correlation is given by:

$$\rho_{ij} = \frac{\mathbb{E}(\mathbf{x}_i - \mu_i) \cdot \mathbb{E}(\mathbf{x}_j - \mu_j)}{\sigma_i \sigma_j} \qquad \text{for} \quad i, j \in \mathcal{D} \tag{2}$$

The diagonal elements of the covariance matrix $\Sigma_{ii} \in \mathbf{\Sigma}$ are the variances of the individual feature vectors, while the off-diagonal elements $\Sigma_{ij} \in \mathbf{\Sigma}$ for $i \neq j$ measure the relationship between feature vectors i and j in the dataset $\mathcal{D}$.

3.1 Data reduction problem

Suppose we wanted to reduce the dimensionality of the feature vectors $\mathbf{x} \in \mathcal{D}$ from m to k dimensions, where $k < m$, i.e., we want to transform $\mathbf{x}_i \in \mathbb{R}^m \to \mathbf{y}_i \in \mathbb{R}^k$. We may want to do this for a variety of reasons, for example, to visualize the data in 2 or 3 dimensions, or to reduce the computational complexity of a machine learning algorithm. To make this possible, we construct (somehow) a projection matrix $\mathbf{P}$ such that:

$$\mathbf{y}_i = \mathbf{P}\mathbf{x}_i \quad i = 1, 2, \ldots, n \tag{3}$$

The projection matrix $\mathbf{P}$ will be a $k \times m$ matrix composed of some transform vectors. The open question is what are the best transform vectors to use? The answer is to use the eigenvectors of the covariance matrix $\mathbf{\Sigma}$. However, we build our transformation matrix using the eigenvectors gives the reduced feature vector corresponding to $x \in \mathcal{D}$ is given by:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{bmatrix} - \; \hat{\mathbf{v}}_1^T \; - \\ - \; \hat{\mathbf{v}}_2^T \; - \\ \vdots \\ - \; \hat{\mathbf{v}}_k^T \; - \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \tag{4}$$

where $\hat{\mathbf{v}}_\star^T$ denotes the transpose of the scaled eigenvector $\hat{\mathbf{v}}_\star$ of the covariance matrix $\mathbf{\Sigma}$. Thus, each component of the reduced vector $\mathbf{y}$ corresponding to $\mathbf{x} \in \mathcal{D}$ is given by:

$$y_j = \hat{\mathbf{v}}_j^T \cdot \mathbf{x} \quad j = 1, 2, \dots, k \tag{5}$$

The open question is which k-vectors should we use to build the transformation matrix $\mathbf{P}$?

Minimize reconstruction error

Fill me in.

4   Summary and Conclusions

In this lecture we discussed the singular value decomposition (SVD) of data and systems, and a method for reducing the dimensionality of data called principal component analysis (PCA). Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) are powerful techniques in linear algebra and data analysis. SVD is a matrix factorization method that decomposes a matrix $\mathbf{A}$ into the product $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{m \times m}$ are orthogonal matrices, i.e., $\mathbf{U} \cdot \mathbf{U}^T = \mathbf{I}$ and $\mathbf{\Sigma} \in \mathbb{R}^{n \times m}$ is a diagonal matrix containing the singular values $\sigma_i = \Sigma_{ii}$. This decomposition has important applications in dimensionality reduction and data compression. On the other hand, principle component analysis (PCA), closely related to SVD, is a method for reducing the dimensionality of multivariate data while preserving as much variance as possible (or alternatively minimize the reconstruction error). It works by identifying principal components, which are orthogonal directions in the data space that capture the most variation (or minimize the construction error). Both SVD and PCA are widely used in various fields, including machine learning, image processing, and gene expression analysis, for tasks such as feature extraction, noise reduction, and data visualization.

References

1. Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci U S A. 2000;97(18):10101–10106. doi:10.1073/pnas.97.18.10101.

2. Alter O, Golub GH. Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. Proceedings of the National Academy of Sciences. 2006;103(32):11828–11833. doi:10.1073/pnas.0604756103.

3. Famili I, Palsson BO. Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices. J Theor Biol. 2003;224(1):87–96. doi:10.1016/s0022-5193(03)00146-2.

4. Moore C. Application of Singular Value Decomposition to the Design, Analysis, and Control of Industrial Processes. In: 1986 American Control Conference; 1986. p. 643–650.

5. Sastry CS, Rawat S, Pujari AK, Gulati VP. Network traffic analysis using singular value decomposition and multiscale transforms. Information Sciences. 2007;177(23):5275–5291. doi:https://doi.org/10.1016/j.ins.2006.07.007.

6. Palacios P, Castro A, Azurdia-Meza C, Estevez C. SVD detection analysis in cognitive mobile radio networks. In: 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN); 2017. p. 222–224.

7. Giuliani A. The application of principal component analysis to drug discovery and biomedical data. Drug Discovery Today. 2017;22(7):1069–1076. doi:https://doi.org/10.1016/j.drudis.2017.01.005.

8. Stewart GW. On the Early History of the Singular Value Decomposition. SIAM Review. 1993;35(4):551–566. doi:10.1137/1035134.