CHEME 5820: Machine Learning for Engineers **Spring 2025**

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: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$, 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 $\mathbf{A} \in \mathbb{R}^{m\times n}$. The SVD of $\mathbf{A}$ is a factorization of the form: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$, where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{m\times m}$ are orthogonal matrices, i.e., $\mathbf{U}\cdot\mathbf{U}^{\top} = \mathbf{I}$ 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 σ_i are the singular values (ordered from largest to smallest) of the 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, i.e., a mode of the original matrix $\mathbf{A}$, with elements:

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

where the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ denote the left (right) singular vectors, respectively, of the matrix $\mathbf{A}$.

Singular value decomposuition is a special sort of eigendecomposition, thus, we could (theorectially) use QR-itereation to compute the SVD. However, what matrix do we use? The columns of $\mathbf{U}$ (left-singular vectors) are eigenvectors of $\mathbf{A}\mathbf{A}^\top$, while the columns of $\mathbf{V}$ are eigenvectors of $\mathbf{A}^\top\mathbf{A}$. Finally, the singular values σ_i are the square roots of the eigenvalues λ_i, i.e., $\sigma_i = \sqrt{\lambda_i}$ of the matrix product(s) $\mathbf{A}\mathbf{A}^\top$ or $\mathbf{A}^\top\mathbf{A}$. Thus, there is a direct relationship between the eigenvalue decomposition of a matrix $\mathbf{A}$ and its singular value decomposition.

3 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a widely used method for reducing the dimensionality of data by projecting it onto a lower-dimensional subspace, see the PCA tutorial by Shlens (9). 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. However, this is the end of the story. Let's start at the beginning, and see how we get there.

3.1 Data reduction problem

Suppose we have a dataset $\mathcal{D} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ where each $\mathbf{x}_i \in \mathbb{R}^m$ is an m-dimensional feature vector. Further, suppose we wanted to reduce the dimensionality of the feature vectors $\mathbf{x} \in \mathcal{D}$ from m to k dimensions, where $k \ll m$, i.e., we wanted to transform $\mathbf{x}_i \in \mathbb{R}^m \rightarrow \mathbf{y}_i \in \mathbb{R}^k$. The reduced-order transformed feature vector $\mathbf{y}_i$ is called a composite feature, as it is a linear combination of the original features. 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:

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

The projection matrix $\mathbf{P}$ will be a $k \times m$ matrix composed of some transform vectors ϕ, that have special properties. First, the k transformation vectors $\phi_1, \phi_2, \dots, \phi_k \in \mathbb{R}^m$, i.e.,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix} = \begin{bmatrix} — \phi_1^\top — \\ — \phi_2^\top — \\ \vdots \\ — \phi_k^\top — \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \tag{2}$$

must be orthonormal, i.e., $\langle \phi_i, \phi_j \rangle = \delta_{ij}$, where δ_{ij} is the Kronecker delta function, i.e., $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, and $\langle \star, \star \rangle$ is the inner product of two vectors. Next, by convention, the transformation vectors are scaled such that $\|\phi_j\| = 1$. However, we know that $||\phi_j|| = \sqrt{\langle \phi_j, \phi_j \rangle} = 1$, by default, if we choose an orthonormal set of transformation vectors.

The open question is what are the best transformation vectors to use? There are two ways to think about this question. The first is to select the k-vectors of that capture the most variance in the data. The second (which I find more intuitive) is to select the k-vectors that minimize the reconstruction error, i.e., we want the reduced composite features to be the best possible approximation of the original data set. These two approaches are equivalent, thus, let's explore the second approach.

Minimize reconstruction error

When talking about PCA and the reconstruction problem we typically assume mean-centered data, i.e., every feature vector $\mathbf{x}_i \in \mathcal{D}$ has been centered by subtracting the mean and dividing by the standard deviation of each feature vector, i.e., $\mathbf{x}_i = \frac{\mathbf{x}_i - \bar{\mathbf{x}}}{\sigma_{\mathbf{x}}}$.

, i.e., every feature vector $\mathbf{x}_i \in \mathcal{D}$ has been centered by subtracting the mean of each feature vector, i.e., $\mathbf{x}_i = \mathbf{x}_i - \bar{\mathbf{x}}$.

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 $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$, where $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{m\times m}$ are orthogonal matrices, i.e., $\mathbf{U}\cdot\mathbf{U}^{\top} = \mathbf{I}$ 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.

9. Shlens J. A Tutorial on Principal Component Analysis; 2014. Available from: `https://arxiv.org/abs/1404.1100`.