(3) Principal component analysis

Study questions

1. Name a potential use of PCA in data analysis.
2. Write down the definition of the covariance matrix of a random vector.
3. What do the elements of a random vector's covariance matrix encode?
4. Given a data matrix $Y \in \mathbb{R}^{m\times n}$ comprising n samples of an m-dimensional random vector y, write down the unbiased estimator for the random vector's covariance matrix.
5. Write down the definition of an eigenvector and an eigenvalue of a square matrix.
6. Explain how eigenvalues and eigenvectors of a square matrix can be computed.
7. Write down the definitions of symmetric, diagonal, and orthogonal matrices.
8. What is the inverse of an orthogonal matrix?
9. What can be said about the eigenvalues and eigenvectors of symmetric matrices?
10. State the theorem on the orthonormal decomposition of a symmetric matrix.
11. Write down the definition of a linear combination of vectors.
12. Write down the definition of the linearly independence of a set of vectors.
13. Write down the definition of a basis of a vector space.
14. Write down the definition of a vector basis expansion and of vector coordinates.
15. Write down the definition of an orthonormal basis of $\mathbb{R}^m$.
16. Write down the canonical basis of $\mathbb{R}^m$.
17. For $y \in \mathbb{R}^m$ and a basis B, write down the orthonormal expansion of y in terms of B.
18. State the change of basis and vector coordinate transform theorem.
19. Write down the definition of PCA. What is the relevance of the matrix of eigenvectors of the data random vector's covariance matrix?
20. What can be said about a PCA-coordinate transformed random vector $\tilde{y} = Q^T y$?
21. Write down the definition of the singular value decomposition of a matrix $X \in \mathbb{R}^{n\times m}$.
22. Let $X = USV^T$ be the SVD of a matrix X. What can be said about the entries of U, S and V in relation to the matrix products XX^T and X^TX?
23. Explain how to compute a PCA for a data matrix $Y \in \mathbb{R}^{m\times n}$ by means of SVD.

Exercises

1. Specify the covariance matrix parameter of a 5-dimensional Gaussian random vector, sample this vector 30 times, and compute the PCA of the resulting 5×30-dimensional data matrix by means of eigenanalysis. Visualize and document your results.
2. Compute the SVD of the same data matrix and evaluate how well the original data matrix can be reconstructed after setting a subset of the singular values to zero. Visualize and document your results.