SVD: A Summary

1. Singular Value Decomposition: $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_i \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

Here, $\mathbf{U}, \mathbf{V}$ have orthonormal columns and Σ is diagonal.

2. Matrix Norms

- Frobenius: $\|\mathbf{A}\|_F = \sqrt{\sum a_{ij}^2} = \sqrt{\sum \sigma_i^2(\mathbf{A})}$
- Spectral: $\|\mathbf{A}\|_2 = \max_{\mathbf{q}:\|\mathbf{q}\|_2=1} \|\mathbf{A}\mathbf{q}\|_2 = \sigma_{\max}(\mathbf{A}) = \sigma_1(\mathbf{A})$
- Nuclear: $\|\mathbf{A}\|_* = \sum \sigma_i(\mathbf{A})$

3. Low Rank Matrix Approximation

The best rank-k approximation of a matrix $\mathbf{A}$ (under both the Frobenius norm and the spectral norm) is $\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$.

4. Principal Component Analysis (PCA)

$$\widetilde{\mathbf{X}} = \mathbf{X} - \bar{\mathbf{x}} = \mathbf{U}\Sigma\mathbf{V}^T$$

Things to keep in mind:

1. The rows of $\mathbf{X}$ represent the given data points, while those of $\widetilde{\mathbf{X}}$ represent centered data.
2. $\bar{\mathbf{x}}$ is the center of the data set, which always lies on the best-fit k-dimensional subspace (that minimizes the total squared orthogonal error).
3. $\mathbf{V}(:, 1:k)$ is an orthonormal basis for the best-fit k-dimensional subspace.
4. The rows of $\widetilde{\mathbf{X}}_k = \mathbf{U}(:, 1:k)\Sigma(1:k, 1:k)\mathbf{V}(:, 1:k)^T$ represent the coordinates of the projections of the centered data onto the best-fit subspace.
5. The rows of $\mathbf{U}(:, 1:k)\Sigma(1:k, 1:k)$, which also equals $\widetilde{\mathbf{X}}\mathbf{V}(:, 1:k)$, are called the top k principal components of the data, being the coordinates of the projections on the best-fit subspace relative to the basis $\mathbf{V}(:, 1:k)$.
6. The right singular vectors (i.e., columns of $\mathbf{V}$) are the principal directions in the data along which the variance of the projections onto $\mathbf{v}_j$ is as large as possible (and equals the corresponding singular value squared, σ_j^2)
7. The number of nonzero singular values is the matrix rank of $\widetilde{\mathbf{X}}$, while the number of “dominant” singular values is the “effective” rank of $\widetilde{\mathbf{X}}$.

In sum, PCA finds in a given data set low dimensional subspaces that

- *minimize* the total squared orthogonal error; and
- *maximize* the variances of the projections; and
- *preserve* the pairwise distances of the points in the data set as closely as possible.

Applications of SVD

- Low-rank matrix approximation
- Subspace fitting
- Data compression (including denoising, dimensionality reduction, visualization)
- Much more: computing matrix pseudoinverse, solving redundant linear systems, etc.