A Survey of Dimensionality Reduction Techniques

RACHEL LEVANGER

This note set gives a short summary of a collection of commonly-used linear and non-linear dimensionality reduction techniques.

Contents

1 Linear Dimensionality Reduction (LDR) Techniques

1.1 Principal Component Analysis (PCA)

PCA is a tool used to describe a data set in terms of its directions of highest covariance. Given a data set $X = \{x^{(i)}\}_{i=1}^m \in \mathbb{R}^n$ with mean $\mu_X = (\mu_1, ..., \mu_n)$ and $\mu_k = (1/m)\sum_{i=1}^m x_k^{(i)}$, the general sequence of steps is as follows:

(1) Compute the covariance matrix Σ associated to the data set X, where $\Sigma_{ij} = \mathrm{cov}(x^{(i)}, x^{(j)}) = (1/n)\sum_{k=1}^n (x_k^{(i)} - \mu_i)(x_k^{(j)} - \mu_j)$.

(2) Diagonalize Σ via the matrices Φ and Λ, where Φ is orthogonal and Λ is diagonal. The matrix Φ has columns that form the orthonormal basis of eigenvectors of Σ, Λ gives the corresponding eigenvalues, and the equation $\Sigma = \Phi\Lambda\Phi^T$ holds. This is always possible since Σ is real and symmetric.

(3) The magnitude of the eigenvalues in Λ provide insight into the number of dimensions along which the data is organized.

(source: https://stat.duke.edu/ sayan/Sta613/2015/lec/IDAPILecture15.pdf)

rachel@math.rutgers.edu