Dimensionality Reduction

Covariance

Covariance is metric used to determine the variance across two dimensions. This is given by

$$cov(X,Y) = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{n-1}$$

The sign of the covariance determines how the two values change w.r.t to changes in each other.

The covariance matrix shows the covariances of all points in a matrix format, where the matrix dimensions are square and the value is the number of attributes or dimensions in the dataset.

Principal Component Analysis

PCA is a powerful tool to analyse data by compression without much loss of information.

The main algorithm

1. Obtain the data
2. Subtract the mean from each of the data dimensions, producing a dataset whose collective mean is zero.
3. Calculate the covariance matrix for the dataset
4. Calculate the eigenvectors and eigenvalues of the covariance matrix as unit vectors
5. Select the eigenvector with the highest eigenvalue as the principal component of the dataset, and do this recursively until the k highest eigenvectors are obtained. Smaller the eigenvalue, smaller the loss of information.
6. Construct a feature vector of all the eigenvectors chosen as columns
7. Derive the new dataset as the transpose of this vector and its product with the original dataset.

To get back the old data, just multiply the inverse of the transpose feature vector with the obtained dataset. Since this is an eigenvector column matrix, its inverse is just the transpose. Optionally, add the original mean to get the data you have started with.

Singular Value Decomposition

$We\ factorize\ the\ matrix\ A\ as$

$A\ =\ U\Sigma V^T$

$where\ U\ is\ orthogonal,\ \Sigma\ is\ diagonal\ and\ V\ is\ orthogonal.$

If A is symmetric positive definite, $A = QSQ^T$, where U = V = Q.

Algorithm

1. Find the eigenvectors of A^TA. The eigenvalue’s square root comes in the sigma matrix.
2. Find the eigenvectors of AA^T. The eigenvalue’s square root comes in the sigma matrix.
3. The eigenvectors of 1 form the U matrix and the eigenvectors of 2 form the V^T matrix.