Tools For Identification Of Structure

Dimension reduction

Geometric Properties

- Norms; Euclidean length

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2}.$$

- Inner products

  The *dot product* or *inner product* of two vectors x and y that have the same number of elements n is denoted by $\langle x, y \rangle$ and is defined by

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i y_i.$$

- The *angle* θ between the vectors x and y is defined in terms of the cosine by

$$\cos(\theta) = \frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}}.$$

- Flats

Linear Transformations

A linear transformation on a vector x is performed by multiplication by a matrix: Ax.

Orthogonality

$$Q^{\mathrm{T}}Q = I,$$

If Q is orthogonal, for the vector x, we have

$$\|Qx\| = \|x\|.$$

Gram-Schmidt Orthogonalization

$$\begin{aligned}
\tilde{x}_1 &= \frac{x_1}{\|x_1\|_2}, \\
\tilde{x}_2 &= \frac{(x_2 - \tilde{x}_1^{\mathrm{T}} x_2 \tilde{x}_1)}{\|x_2 - \tilde{x}_1^{\mathrm{T}} x_2 \tilde{x}_1\|_2}.
\end{aligned}$$

Linear Transformations

Geometric properties

Important characteristics of transformations are what they leave *unchanged* (that is, their *invariance properties*).

Another isometric transformation is a *translation*, which for a vector x is just the addition of another vector:

$$\tilde{x} = x + t.$$

A transformation that preserves angles is called an *isotropic transformation.* A transformation that preserves parallel lines is called an *affine transformation.* All of these transformations are *linear transformations* Rotations

Linear Transformations

Two major tools in seeking linear structure are rotations and projections of the data matrix X.

Translations using *homogeneous coordinates*.

Measures of Similarity and Dissimilarity

Similarities: Covariances and Correlations

Similarities between Groups of Variables

Dissimilarities: Distances

Properties of Dissimilarities

Effects of Transformations of the Data

Outlying Observations and Collinear Variables

Multidimensional Scaling: Determining Observations that Yield a Given Distance Matrix