1. Distance

* Axioms
  + Symmetry
  + Positivity
  + Minimality
  + Triangle inequality

2. principle component analysis (PCA)

* Not a distance model but represent scalar product between vectors as correlation.
* Matrix background
  + Rank of matrix can be thought of as the dimensionality of the data space spanned by the **columns** (rows).
  + Singular matrix (not full rank) cannot do PCA or factor analysis.
  + Trace of the symmetric matrix (e.g., correlation matrix) is the sum of the diagonal elements or the **sum of the eigenvalues**. Thus, the trace can explain the total variance (not include the covariance) in the matrix.
* Let R to the empirical correlation matrix, and them use the singular value decomposition. . Each column of P is the called principal component. P is called component loading matrix. Actually, .P = EA^1/2. Where E is the eigenvector matrix of R. And A is the diagonal matrix with the eigenvalue. Then . The principal components are orthogonal and the corresponding eigenvalue is its variance.
* The sum of the eigenvalue must be equal to the dimension of the R, which is the correlation matrix.
* Approximate the correlation with the reduced number of components.

3. Metric MDS

* Give the distance-like matrix (“dissimilarity”), we want to construct the configuration matrix so that the distance in the space are **linearly** related to “dissimilarity”.
  + Configuration matrix (X) is n times r: n is the points in the space and r is the dimension of the space
  + We have a distance between to vector and we want to transfer this distance into a scalar.
    - For example, using the angle of two vector to represent the distance between two vectors.
    - Cosine law:
    - Torgerson (1957)
      * Given the n times n distance matrix (that satisfied matrix axiom), double-centering. The new distance is linearly related to the original distance
      * Do the PCA to factorize it
      * Additive constant problem: add bigger number need more dimension to distinguish the distances between the points

4. Non-metric MDS 1: The Kruskal's Algorithm

* + The monotonical least square solution 🡺 local minimum solution (stochastic starting points)
  + Input:
    - the proximity from the dissimilarity data, which provide the order of pair distance
    - the lower dimensional configuration (e.g., Torgerson result)
  + steps:
    - derived the Euclidean distance from the lower dimension configuration
    - reset the distance estimation based on the order from the dissimilarity
  + two approach:
    - primary: free
    - secondary: constrain
  + measurement of model fit:
    - stress indexes (the numerator term is sometimes called “raw stress”):

Or, the recommended stress index:

* + Notes:
    - The dimension of the solution space is determined in advance.
    - Only ordinal information in proximities maintained, so relative more data required. Rule of thumb: 7+ stimuli per dimension.
    - To interpret the solution better: regression of “attributes” into the space, then plot the regression coefficients.

5. Non-metric MDS 2: Samcof Algorithm

* Fit the MSD configuration by majorization
  + Replace iteratively the original complicated function by an auxiliary function , where is some fixed value. is simpler to minimize, always bigger or equal to , and touch the surface of at the supporting point .
* General method
  + minimize stress w.r.t a matrix
  + is the weight, can be used to adjust the influence from the missing data
  + Taking the partial derivate of stress w.r.t a matrix X