Multivariate Analysis of Variance (MANOVA)

Finds the mean difference between the groups. Can be tested by Hotellings T (two groups) and Manova (several groups). Hotellings t test is an extension of students t test to the multivariate case and manova is an extension of anova but it uses the variance-covariance between variables in testing the statistical significance of the mean differences. Both rely on the normal distribution assumption. Null hypothesis is no difference in group means.

Manova helps to answer

   Do changes in the independent variable(s) have significant effects on the dependent variables?
   What are the relationships among the dependent variables?
   What are the relationships among the independent variables?

Principal Component Analysis (PCA)

Aims to reduce the number of variables (with better interpretation of new variables and without losing information – unsupervised learning). Dimension reduction technique

In PCA we consider new variables that are linear combinations of the original variables. The aim is to preserve the structure of the data through preserving variance. The idea is to replace the original variables with the new standardized and uncorrelated linear combinations without losing variance or information. Principal components analysis is sensitive to changes in measurement units.

High variation = high information content. First principal component has the highest variance etc. we can see this from the scree plot

Canonical Correlation Analysis (CCA)

Considers simple linear dependencies between two sets of variables. The aim is to replace the original variables with a new smaller set of variables that are standardized and uncorrelated linear combinations of the original variables without losing much information. The new variables are called canonical variables. Dimension reduction technique

A typical use for canonical correlation is to take two sets of variables and see what is common amongst the two sets.

If we have two vectors $X = (X_1, ..., X_n)$ and $Y = (Y_1, ..., Y_m)$, then canonical-correlation analysis will find linear combinations of the X_i and Y_j which have maximum correlation with each other

High correlation = high information content.

Factor Analysis (FA)

Aims to reduce the number of variables (with better interpretation of new variables and without losing information – unsupervised learning) Dimension reduction technique

Estimate how many factors explain the data the best. For example, it is possible that variations in say six observed variables mainly reflect the variations in two unobserved (underlying or latent) variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modelled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.

The variation in the measurements is due to the variation in the common factors. These common factors then carry the information.

Independent Component Analysis (ICA)

Aims to reduce the number of variables (with better interpretation of new variables and without losing information – unsupervised learning) Dimension reduction technique

The new variables are standardized and uncorrelated linear combinations of the old variables. First component has highest kurtosis etc. The idea is to replace the original variables with the new standardized and uncorrelated linear combinations without losing information.

(Just a few are usually used in the model as not every variable holds information, the others are just noise.)
Often used in signal analysis and image analysis to separate independent sources.

High/low kurtosis = high information

Linear and Quadratic Discriminant Analysis (LDA,QDA)

Based on multivariate observations with known class labels, this develops simple discrimination rules for a classification of new individuals with known class labels (supervised learning)

Based on training data with variables that have known class labels a prediction (separation rule) is developed that can be used to assign a new variable to a class. The accuracy and other properties can be tested in the test data set. There is a discriminant function that does the separating.

In knn there is no discriminant function. For a new observation find the k nearest observations and classify the new observation to the class that has a majority among the k nearest neighbors.

Cluster Analysis

Aims to group the observations into natural subsets or clusters. Within the clusters, the observations are as similar as possible. The clusters may then be used to describe a

population and are often arranged into a natural hierarchy. In hierarchical clustering one uses a measure of dissimilarity. Dissimilarity (distance) matrix.

Hierarchical clustering is bottom up. Each observation is a cluster, Find shortest distance (mahalnobis or Euclidian etc) between two clusters and merge them, repeat 2 until the number of clusters is acceptable.

k-means clustering: pick initial value for k, observations are then grouped to clusters with a closest mean vector, then the k clusters mean vectors are clustered again and this continues until there are no changes anymore.