

Lecture Outline

Feature Selection and Extraction

Michael J. Watts

<http://mike.watts.net.nz>

- Regression analysis
- Analysis of variance
- Principal component analysis
- Independent component analysis

Regression Analysis

- Relates one variable to another
 - Dependent / independent
- Finds the “line of best fit”
- Derives a polynomial
- Least squares method
 - Minimises the square error
 - Linear function

Analysis of Variance

- ANOVA
- Collection of techniques
- Way of determining if there is a significant difference
 - Similar to a *t*-test
- MANOVA
 - Difference between multiple variables

Feature Extraction

- Finding relevant variables in the data
 - Significant variables
- Reduces dimensionality
 - Helps visualisation
 - Simplifies analysis
 - Eliminates noise
- Variables that affect the process
 - Others are secondary/unimportant

Feature Extraction

- Correlation between variables
 - Dependence / independence
- A good set of features is
 - Sufficient
 - Includes important features
 - Not redundant
 - Minimal

Principal Component Analysis

- PCA
- Feature extraction method
- Combines correlated variables
 - Linear combination
- Results are uncorrelated
 - Principal components
- Derived from variance of each variable

Principal Component Analysis

- Arranges components by order of variance
 - First is most variant
- Orthogonal
 - Perpendicular to each other
- Variables are combined and extracted
 - A variable ends up in one principal component
 - Always reduces dimensionality

Independent Component Analysis

- ICA
- Finds independent components
 - Assumes variables are linear combinations of the independent components
- ICA is claimed to be more powerful than PCA
- Separates signal from noise
 - Cocktail party problem

Summary

- Statistical methods support feature extraction
- Feature extraction reduces the dimensionality of data
 - Assists analysis
 - Assists visualisation
- Two major feature extraction methods
 - PCA
 - ICA