Dimensionality reduction

Multivariate data

Univariate

- One variable

Bivariate

- Two variables

Multivariate

- Many variables
- Also multidimensinal

Dimensionality reduction

The *curse* of higher dimensions

- Computational ineffectivity.
- Low data density in higher dimensions.
- Problematic visualization, human brain does not easily cope with more than 3D.
- Difficult interpretation.

Principal components analysis

Goal: Find low-dimensional representation of the observations that explain a good fraction of the variation.

- First principal component is a direction that maximizes the variance of the projected data.
- Second PC is orthogonal to the previous one.

Biplot