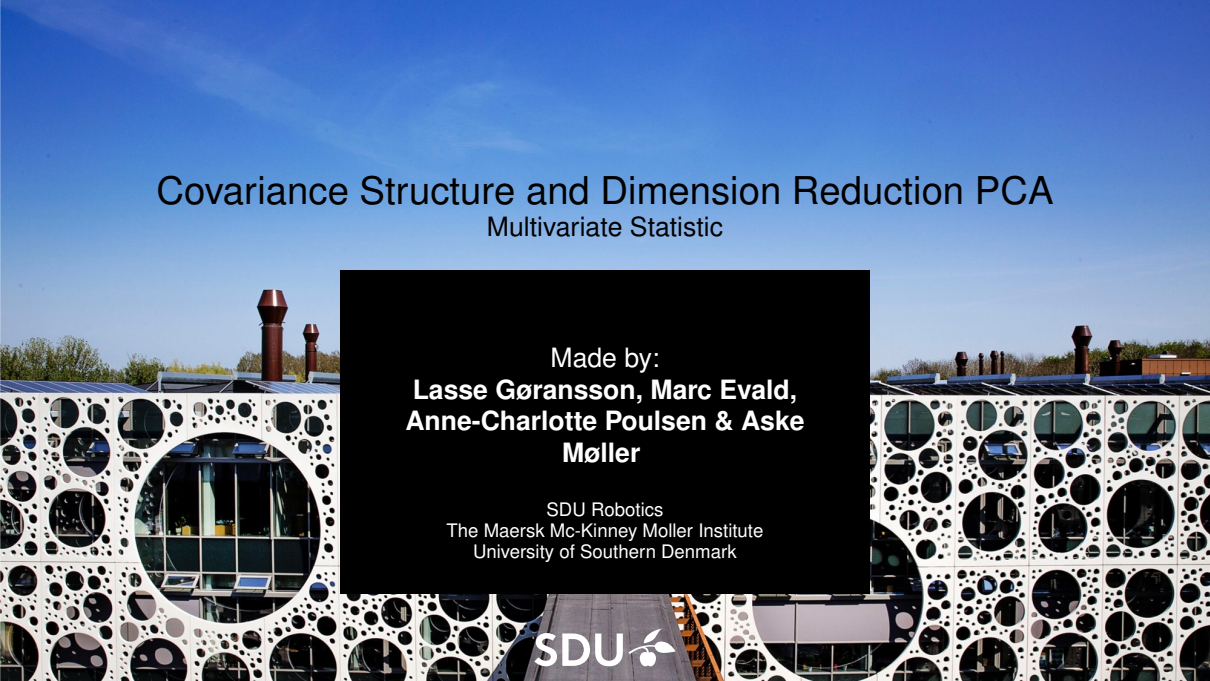
Covariance Structure and Dimension Reduction PCA

Multivariate Statistic

Made by:
Lasse Gøransson, Marc Evald, Anne-Charlotte Poulsen & Aske Møller

SDU Robotics
The Maersk Mc-Kinney Moller Institute
University of Southern Denmark



Agenda

- Motivation
- Principal Components
- Variance
- Sample Based
- Pitfalls

Motivation

- Dimension reduction
- Decorrelate data

Principal Components

$$PC_i \stackrel{def}{=} e_{i1}X_1 + \ldots + e_{ip}X_p = e_i^T X$$

Principal Components

$$PC_i \stackrel{def}{=} e_{i1}X_1 + \ldots + e_{ip}X_p = e_i^T X$$

- Maximize variance

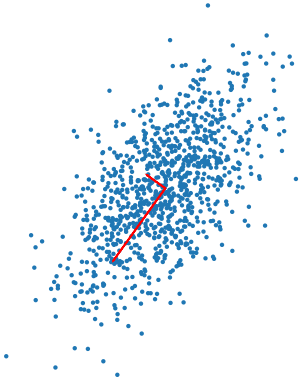
Variance Explained

$$\Sigma = E\Lambda E^T$$

$$PC_i : \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j}$$

Visualization

Visualization



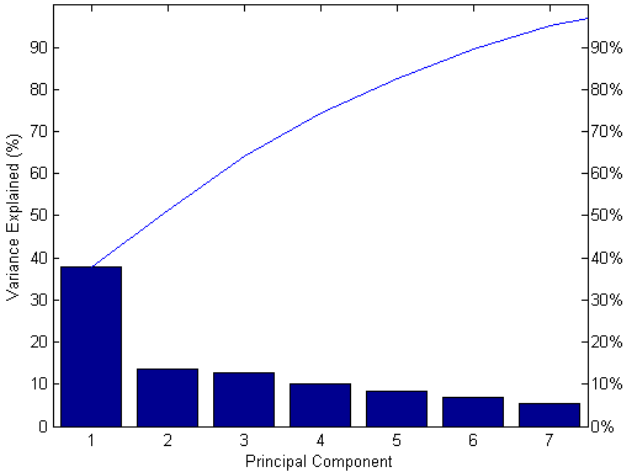
Sample Based

$$X = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix}$$

$$\hat{\Sigma} = S = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T$$

$$\frac{\sum_{i=1}^{r} \hat{\lambda_i}}{\sum_{j=1}^{p} \hat{\lambda_j}}$$

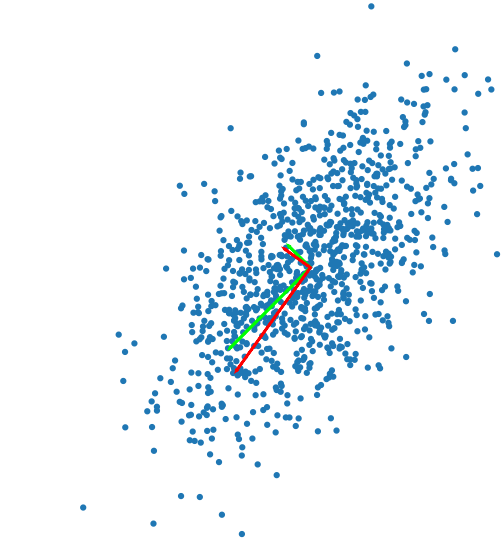
Visualization

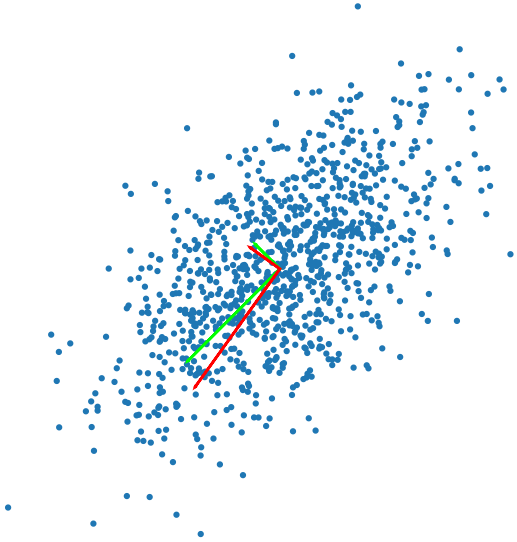


Pitfalls

$$X = \begin{bmatrix} X_{11} & \dots & X_{1p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{bmatrix}$$

Pitfalls





Inference

$$\hat{\lambda_i} \sim N\left(\lambda_i, 2\frac{\lambda_i^2}{n}\right)$$

$$\frac{\hat{\lambda_i}}{1+\sqrt{\frac{2}{n}}z_{\frac{\alpha}{2}}} < \lambda_i < \frac{\hat{\lambda_i}}{1-\sqrt{\frac{2}{n}}z_{\frac{\alpha}{2}}}$$