PCA

data X

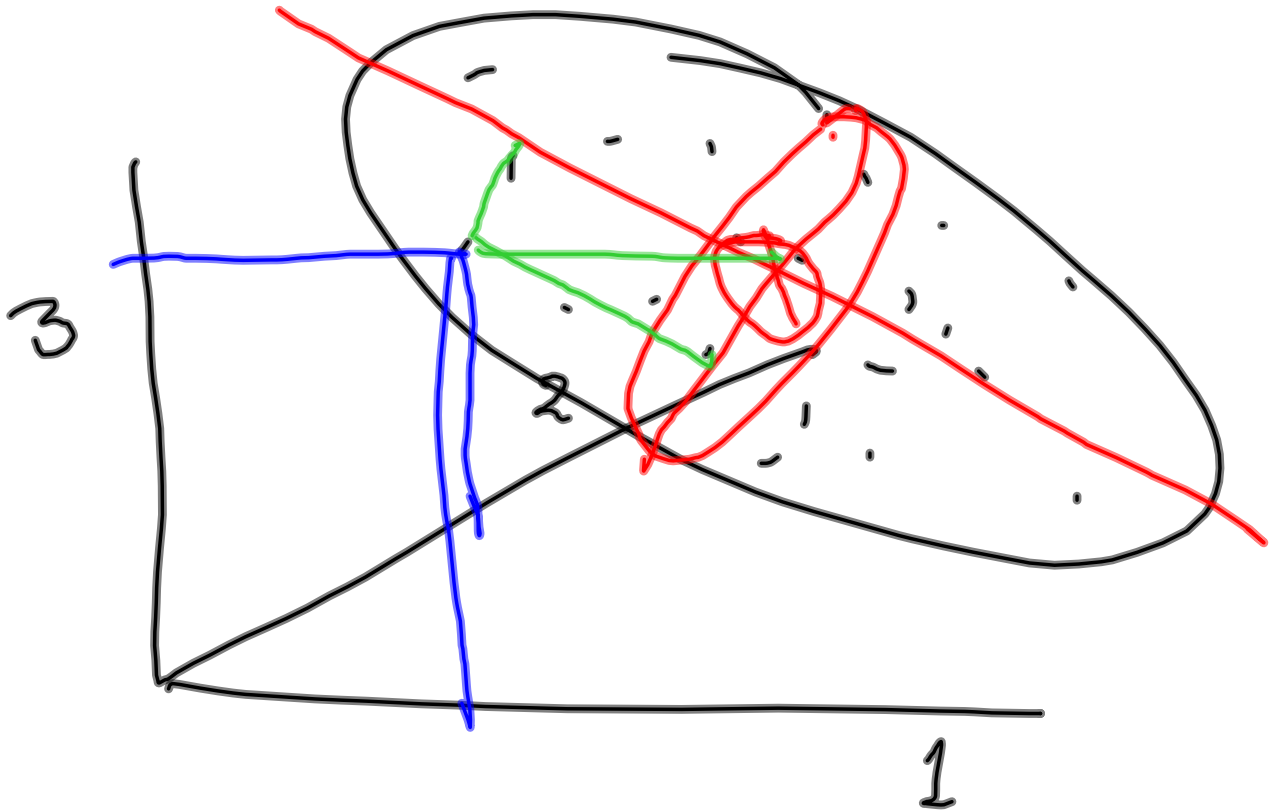
n rows

m columns (variables)

motivation (A) combine variables

motivation (B) decompose observations

transformation of the data (rotation of axes)


3
2
1

Mathematical goal

1. Variance maximize,
   subject to
2. orthogonality

Mathematical Solution

Covariance matrix of m original variables

$$\Sigma_x$$

Eigen decomposition

$$U \Lambda U^T = \Sigma_x$$

General definition of eigvalue λ eigenvector $\underline{v}$

$$B\underline{v} = \lambda \underline{v}$$

$$U \Lambda U^{-1} = B$$

Eigenvalues + Eigenvectors of Σ_x

$$U^T X^T = S^T \qquad X^T = U S^T$$

combining variables where weights are elements of the eigenvector

$S_{n\,m}$ $\Sigma_S = \Lambda$

Original observations are weighted sums of eigenvectors where the scores are the weights

Component scores are weighted sums of the original variables, where eigenvector elements are the weights

Alternative "Space" for viewing data:

1) space of the original m variables (original data)
2) space of the m principal components
3) filtered space of the first k principal components