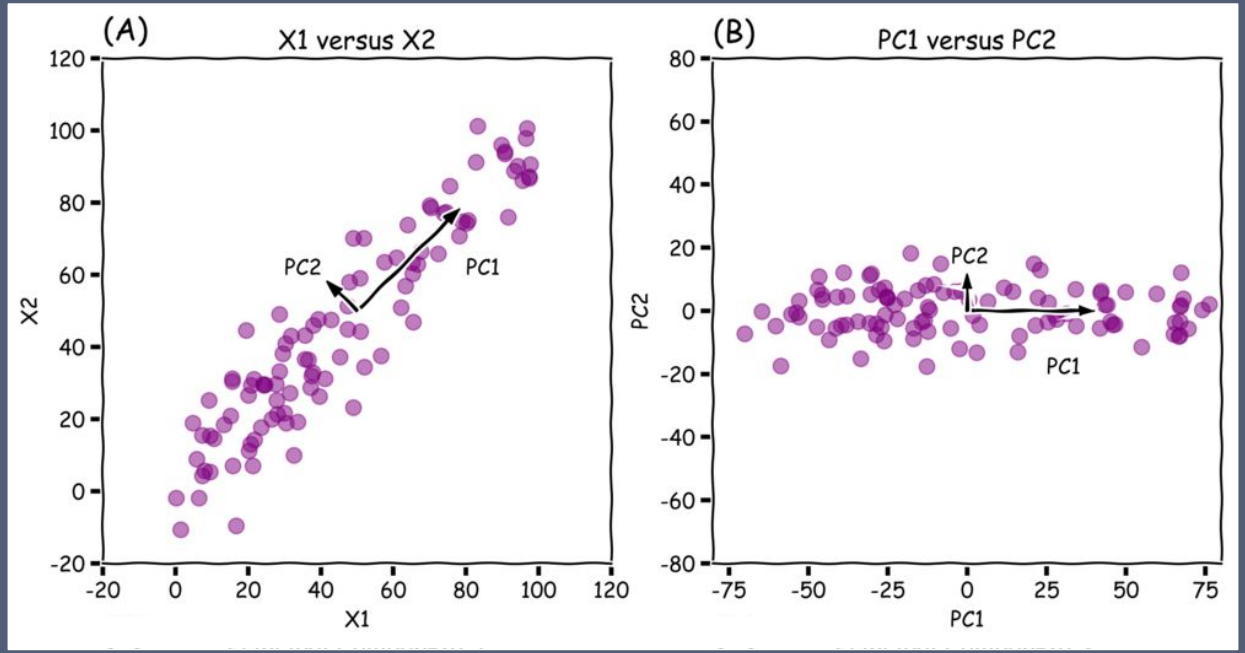
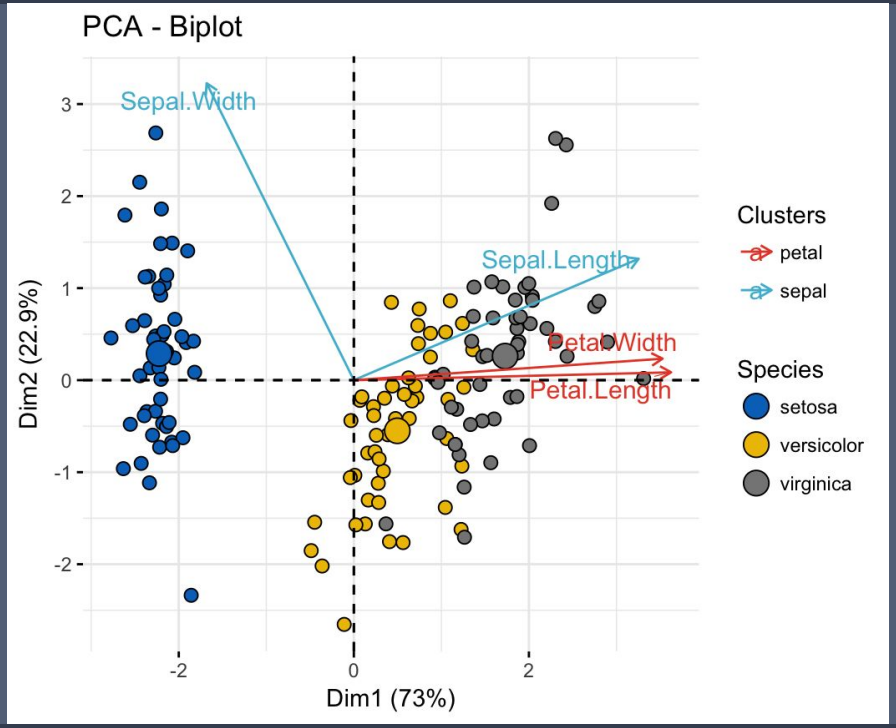
Index Class

1. Correlation
2. Principal component Analysis

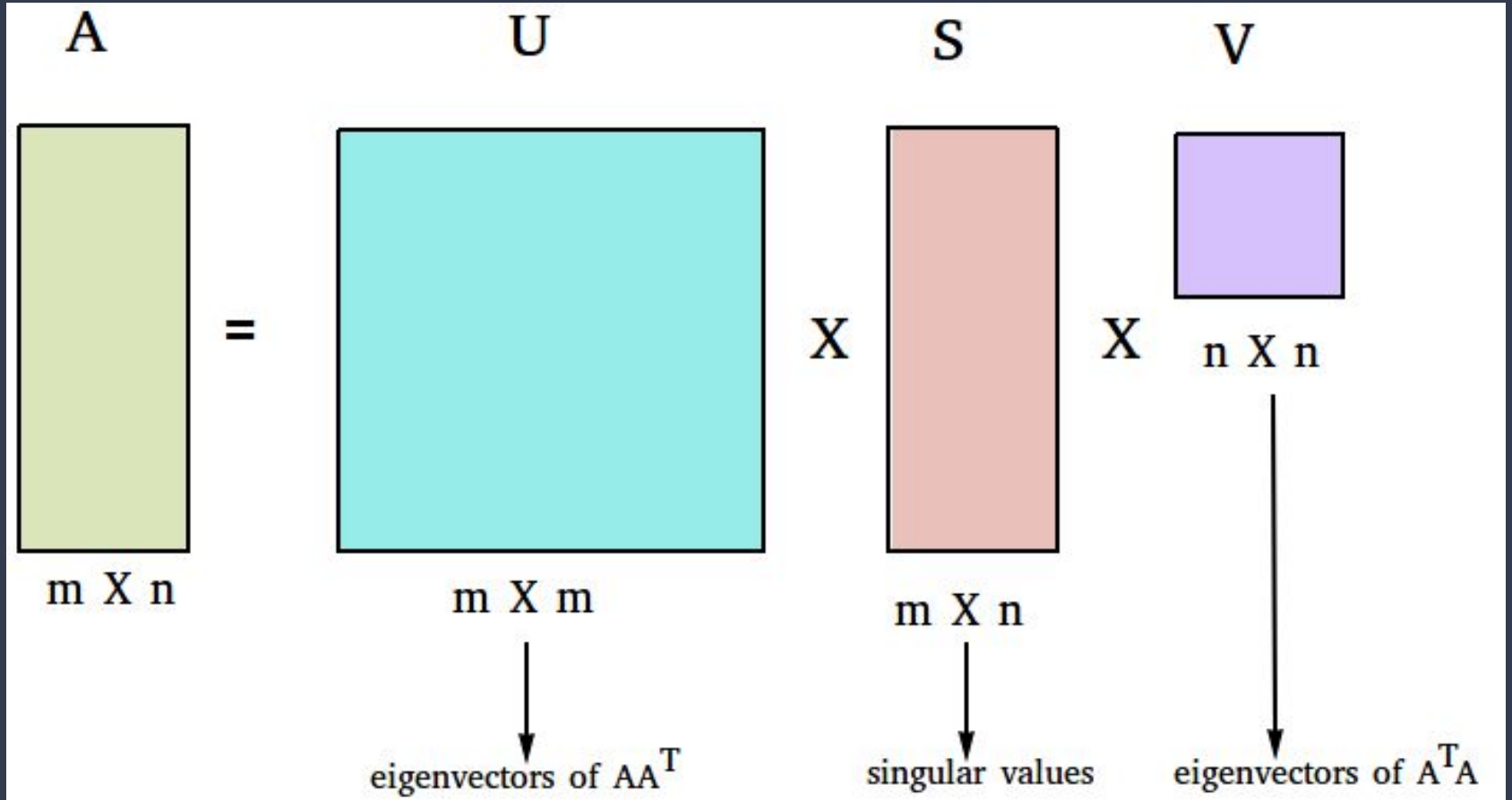
Principal Component Analysis PCA



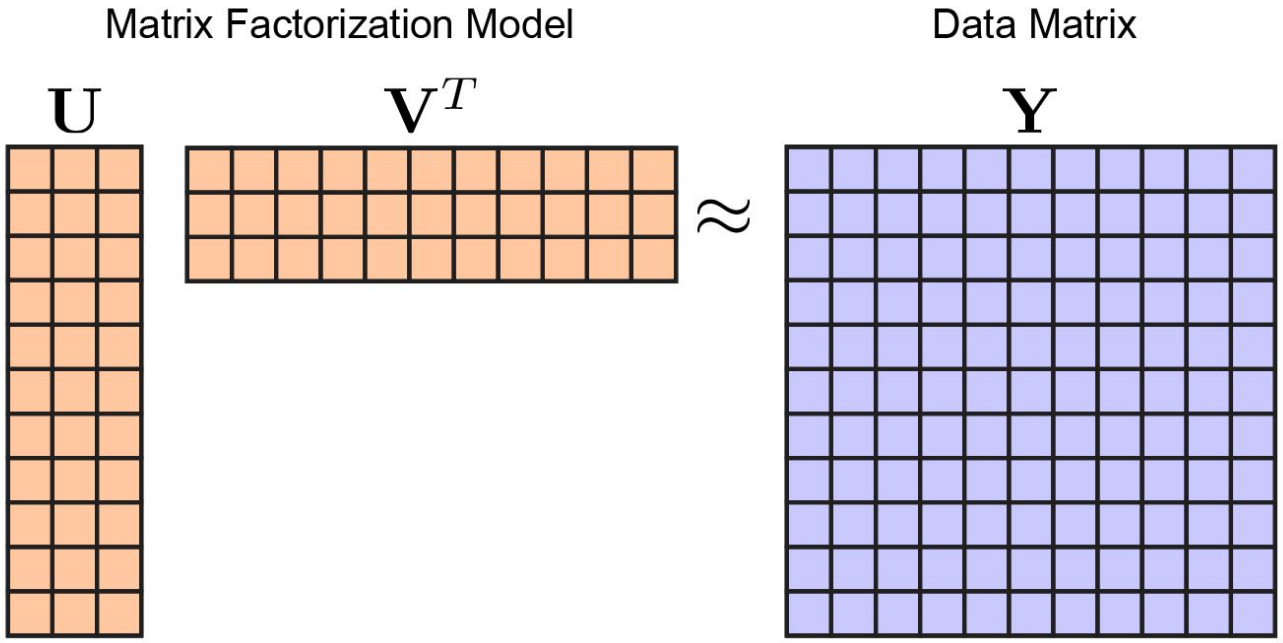
Principal Component Analysis PCA



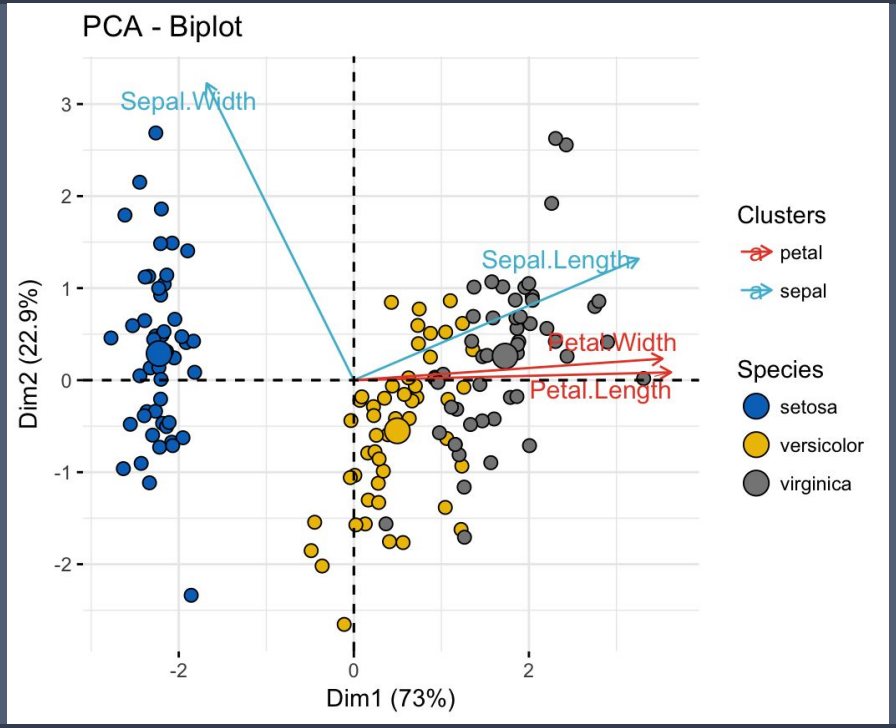
SVD

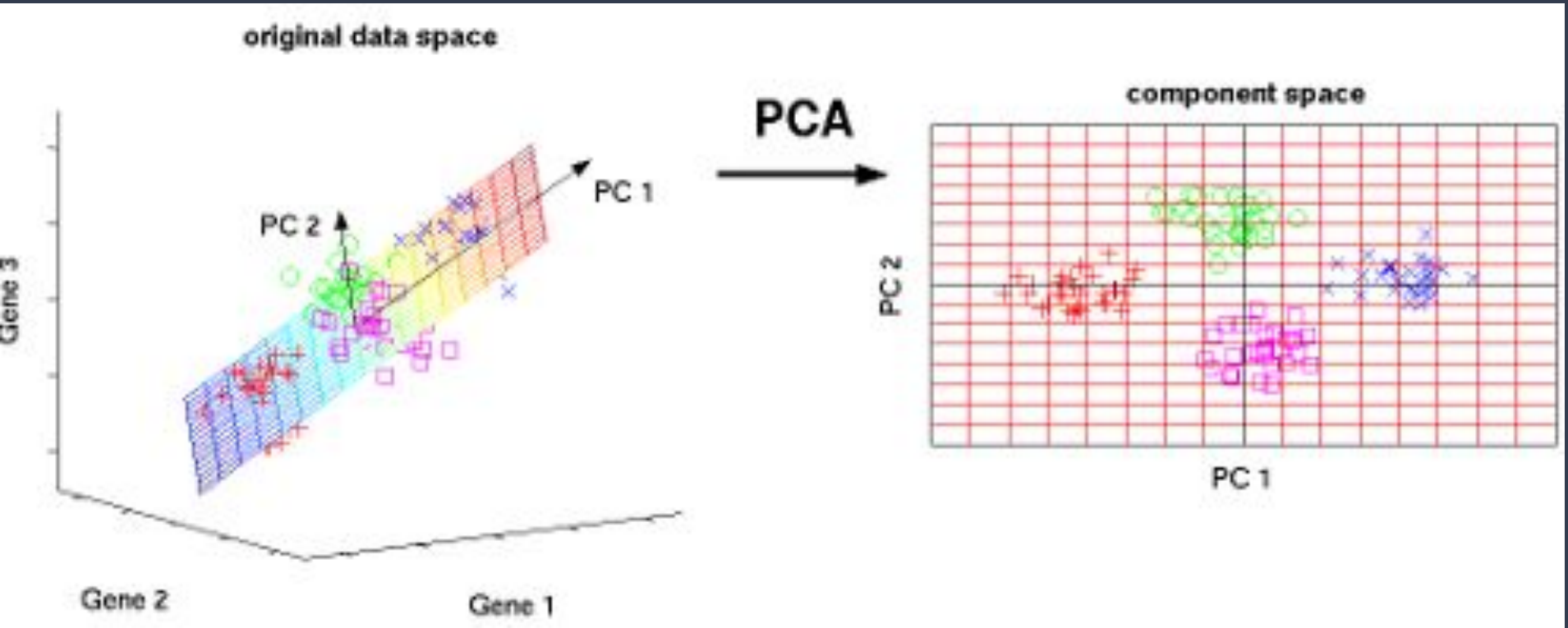

A
U
S
V
=
X
X
m X n
m X m
m X n
n X n
eigenvectors of AA^T
singular values
eigenvectors of A^TA

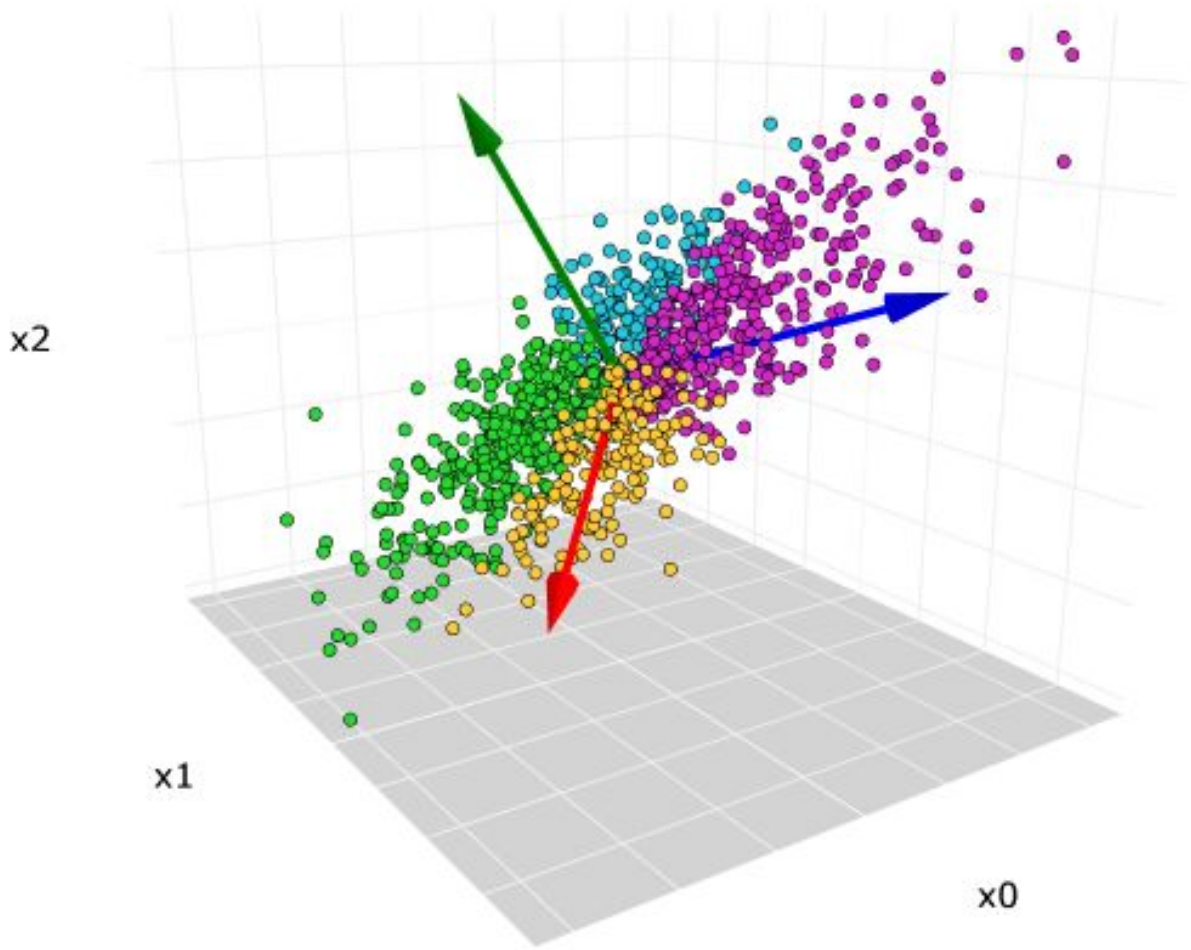
Principal Component Analysis PCA



Principal Component Analysis PCA




original data space
PC 1
PC 2
Gene 3
Gene 2
Gene 1
PCA
component space
PC 2
PC 1


x2
x1
x0

Correlation

Correlation is the covariance normalized

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

http://guessthecorrelation.com/

Principal Component Analysis PCA

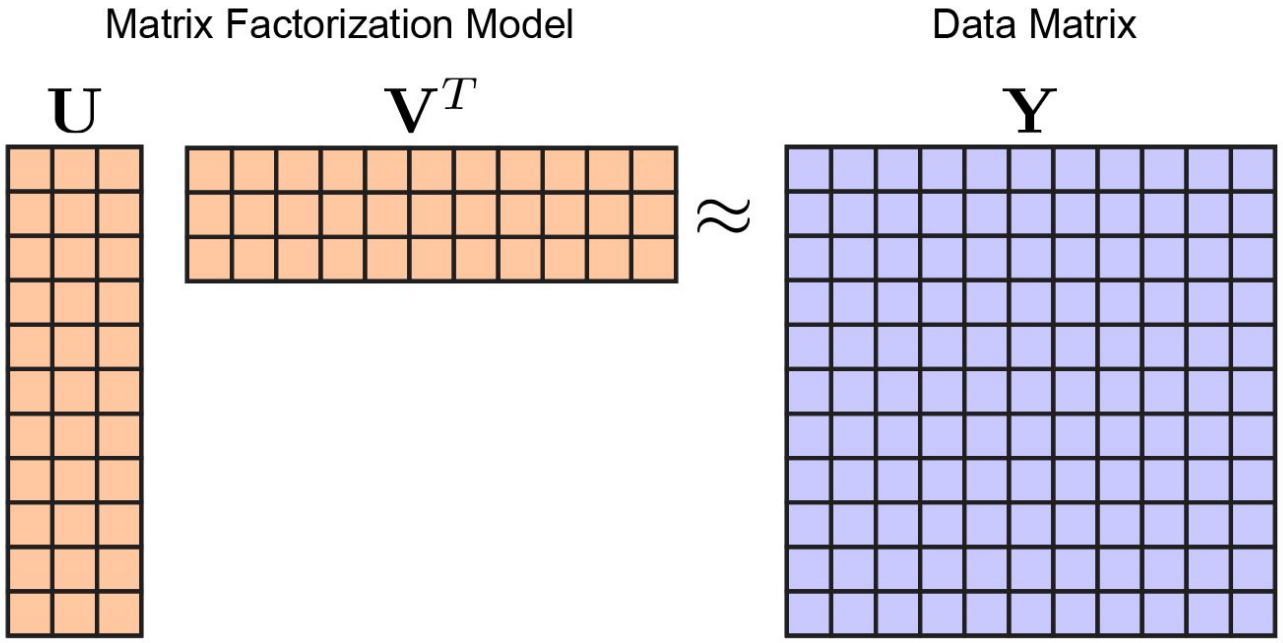
Principal Component Analysis (PCA) is a **dimension-reduction** (Unsupervised algorithm)

Transforms a number of correlated variables into a number of **uncorrelated** variables called **principal components**.

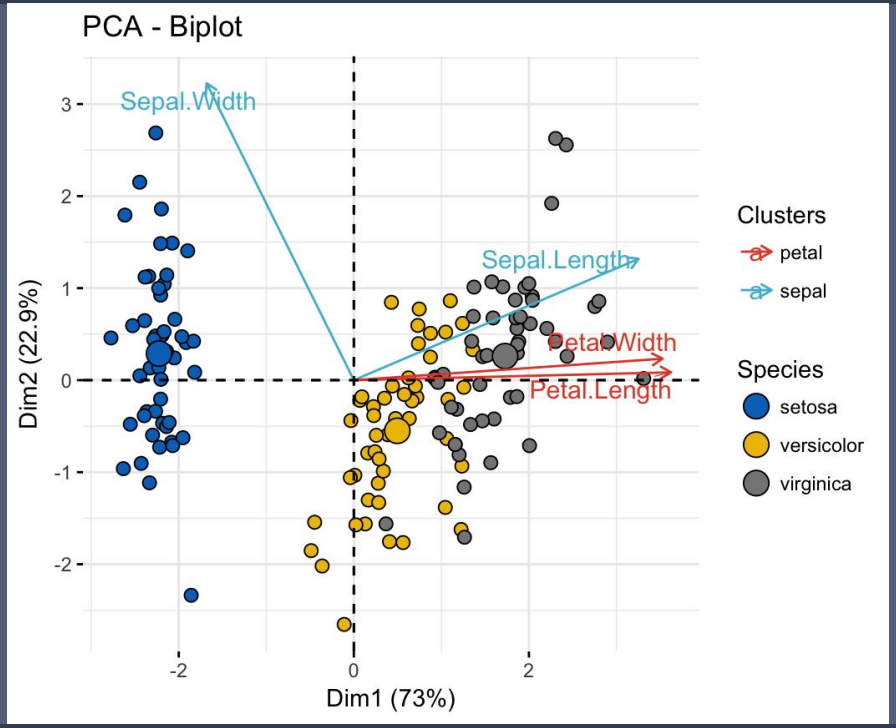
Principal Component Analysis PCA

The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

Principal Component Analysis PCA



Principal Component Analysis PCA



Principal Component Regression

Principal components regression PCR produces the weight matrix W reflecting the covariance structure between the predictor variables.

They use the first components of a PCA to predict the target variable.

Principal Component Regression

Good:

- We have “variables” uncorrelated.
- Avoid overfitting.
- Avoid multicollinearity.

Bad:

- Those components have no relation with the target.