
H2O.ai

Pros and Cons of K-Means Clustering

Pros

- Fast, Scalable Algorithm

Cons

- Choice of k can be tricky
- Euclidean distance not robust
  - Hyperspheres not common
  - Sensitive to correlated measures
  - Sensitive to scaling
  - Sensitive to skewed measures
  - Sensitive to outliers
- Categorical data requires preprocessing
  - Multiple Correspondence Analysis
  - Multi-Dimensional Scaling

Unsupervised Learning:

PRINCIPAL COMPONENTS ANALYSIS (PCA)

Pros and Cons of K-Means Clustering

Pros

- Fast, Scalable Algorithm

Cons

- Choice of k can be tricky
- Euclidean distance not robust
  - Hyperspheres not common
  - Sensitive to correlated measures
  - Sensitive to scaling
  - Sensitive to skewed measures
  - Sensitive to outliers
- Categorical data requires preprocessing
  - Multiple Correspondence Analysis
  - Multi-Dimensional Scaling

H2O.ai