

# 数据科学家直通车 Phase 2 - Week 3 实战课

By: John



BITTIGER

The Lifelong Learning Platform of Silicon Valley

# Outline



- PCA and Matrix Factorization
  - PCA
  - NMF
- Recommender System
  - Item-item collaborative filtering
  - Matrix Factorization
    - NMF
    - UVD

# 1. PCA

Location: 1-PCA\_and\_Matrix\_Factorization/PCA.ipynb



**BITTIGER**

The Lifelong Learning Platform of Silicon Valley

# PCA



- Implement PCA
  - Use `sklearn.decomposition.PCA`
  - Solve eigen value problem using `numpy.linalg`
- Determine number of PCs to keep
  - Percentage of variance explained, e.g. 90%
  - Visualize
- PCA Regression
  - How to determine best number of PCs for model?
    - Cross-validation error
  - Problems with PCA Regression:
    - Loss of interpretation

## 2. NMF

Location: 1-PCA\_and\_Matrix\_Factorization/NMF-Demo.ipynb



**BITTIGER**

The Lifelong Learning Platform of Silicon Valley

# NMF



- Implement NMF
  - Use `sklearn.decomposition.NMF`
- Implement SVD
  - Use `numpy.linalg.svd`
- Compare NMF with SVD for dimension reduction
- Interpretation of NMF using concept (fuzzy clustering)
  - The concepts are clusters
  - Q1: What do the concepts (clusters) mean?
  - Q2: To which concept(s) does each user/document belong?

### 3. Item-Item Collaborative Filter Recommender

Location: 2-Recommender/Item-Item\_CF\_Recommender.ipynb



**BITTIGER**

The Lifelong Learning Platform of Silicon Valley

# Item-Item Collaborative Filter Recommender



- Convert raw data to utility matrix
  - Use Pandas pivot table
  - Use `scipy.sparse.lil_matrix`
- Perform item-item based recommendation



## 4. Matrix Factorization Based Recommender

Location: 2-Recommender/NMF\_UVD\_for\_recommender.ipynb



**BITTIGER**

The Lifelong Learning Platform of Silicon Valley

# Matrix Factorization Based Recommender



- Convert raw data to utility matrix
  - Use `scipy.sparse.lil_matrix`
- Perform Matrix Factorization Based Recommender
  - Use NMF: `sklearn.decomposition.NMF`
  - Use UVD: `sklearn.decomposition.TruncatedSVD`

# Summary



- PCA and Matrix Factorization
  - PCA
  - NMF
- Recommender System
  - Item-item collaborative filtering
  - Matrix Factorization
    - NMF
    - UVD