Plagiarism Detector .net

Dec 30, 2024

Plagiarism Scan Report

Characters:2738	Words:414
Sentences:16	Speak Time: 4 Min

Excluded URL	None

Content Checked for Plagiarism

AIE425 Intelligent Recommender Systems Assignment #3: Dimensionality Reduction methods Student ID: A20000882, Student Name: Maryam Elgohary Qasim 1. 1. Discussion Part 1 PCA with Mean Filling Output: 1: Output: top 10 peers predictions: Top 5 peers predictions: Advantages of PCA mean-filling: 1- Handling missing data 2- Dimensionality reduction helps reducing the complexity of data making it focuses on most important features only 3- Covariance matrix helps getting the relationships between items Disadvantages of PCA mean-filling: 1- There is bias 2- Can have computation complexity on large dataset as it fill every null value with mean and compute based all the dataset Comment: The top 5 peer gets more focused recommendations while 10 peers get a more diverse recommendation. There is bias towards average ratings Part 2 PCA with MLE: Output: Output: Top 10 peer predictions: Top 5 peer predictions: Advantages of PCA MLE: 1- Higher accuracy 2- Handles missing data Disadvantages of PCA MLE: 1- More computation 2- Over fitting risk 3- Requires Data Quality Comment: Top 5 peers are more stable in top 10 peers there are more variations. MLE-based prediction provides a more sophisticated and data-driven approach. Part 3 SVD: Z-matrix: V-matrix: U-matrix: Reconstructed R Matrix: Advantages of SVD: 1- Noise reduction 2- Handle missing data 3- Scalability 4- Works on sparse Disadvantages of SVD: 1. High Computational cost 2. Over fitting 3. Assuming linear relationship Comment : SVD is a powerful technique for collaborative filtering, especially when dealing with large-scale datasets. It is particularly useful when there are latent factors between users and items that can be captured by decomposing the user-item matrix. Comparison of the three methods 2. Conclusion . PCA with Mean Filling is one of the simplest and fastest methods that can be applied on small data or for computational efficiency reasons. However, this approach may cause bias in cases \ PCA using MLE offers a more statistically rigorous approach. It outperforms EM when missing is not random and the data is from a known distribution. However, it requires more computational resources and domain knowledge about the distribution of the data, making it less practical on very large datasets. SVD is particularly effective for PCA with big and sparse datasets. Because of its matrix factorization, it naturally handles missing values by approximating them based on latent factors. While robust for the capture of hidden relationships, this method may be less accurate in cases where the rank approximation is not well aligned with the structure of the data 4

Sources