

MITx 6.86x

## **Machine Learning with Python-From Linear Models to Deep Learning**

Discussion Course **Progress Resources** Dates

A Course / Unit 4. Unsupervised Learning... / Project 4: Collaborative Filtering vi



< Previous

 $\square$  Bookmark this page

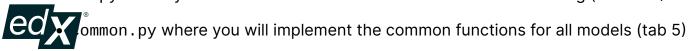
Your task is to build a mixture model for collaborative filtering. You are given a data material ratings made by users where the matrix is extracted from a much larger Netflix database has rated only a small fraction of the movies so the data matrix is only partially filled. The remaining entries of the matrix.

You will use mixtures of Gaussians to solve this problem. The model assumes that each sample from a mixture model. In other words, we have K possible types of users and, it user, we must sample a user type and then the rating profile from the Gaussian distributive. We will use the Expectation Maximization (EM) algorithm to estimate such a mixture observed rating matrix. The EM algorithm proceeds by iteratively assigning (softly) use subsequently re-estimating the Gaussians associated with each type (M-step). Once we can use it to predict values for all the missing entries in the data matrix.

## Setup:

As with the last project, please use Python's **NumPy** numerical library for handling arrayuse **matplotlib** for producing figures and plots.

- Note on software: For all the projects, we will use python 3.6 augmented with the Nu
  the matplotlib plotting toolbox. In this project, we will also use the typing library, whi
  the standard library (no need to install anything).
- 2. Download <u>netflix.tar.gz</u> and untar it in to a working directory. The archive contains th
  - kmeans where we have implemented a baseline using the K-means algorithm
  - naive\_em.py where you will implement a first version of the EM algorithm (tabs 3-
  - em.py where you will build a mixture model for collaborative filtering (tabs 7-8)



• main.py where you will write code to answer the questions for this project

ed X test.py where you will write code to test your implementation of EM for a given to

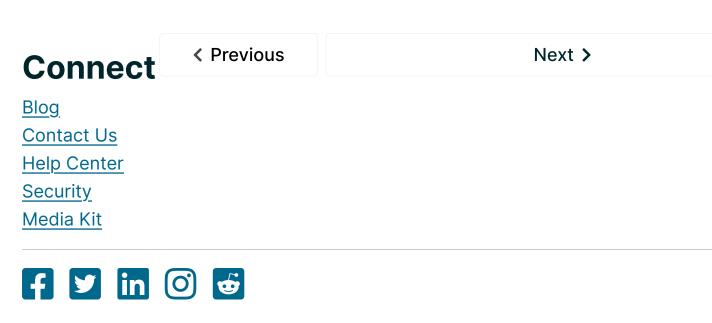
About Additionnally, you are provided with the following data files: Affiliates

edX for Businesstxt a 2D dataset that you will work with in tabs 2-5

Open edX incomplete.txt the netflix dataset with missing entries to be completed Careers • netflix\_complete.txt the netflix dataset with missing entries completed

News

. toot incomplete tyte toot detect to test for you to test your ende against our







© 2023 edX LLC. All rights reserved.

深圳市恒宇博科技有限公司 <u>粤ICP备17044299号-2</u>