



MITx 6.86x

Machine Learning with Python-From Linear Models to Deep Learning[Course](#)[Progress](#)[Dates](#)[Discussion](#)[Resources](#)[Course](#) / [Unit 4. Unsupervised Learning...](#) / [Project 4: Collaborative Filtering vi](#)[< Previous](#)

8. Using the mixture model for collaborative filtering

[Bookmark this page](#)

Project due Apr 26, 2023 08:59 -03 Past due

Reporting log likelihood values on Netflix data

0.0/1.0 point (graded)

Now, run the EM algorithm on the incomplete data matrix from Netflix ratings `netflix` before, please use seeds from `[0, 1, 2, 3, 4]` and report the best log likelihood you achieve for $K = 12$.

This may take on the order of a couple minutes for $K = 12$.

Report the maximum likelihood for each K using seeds `0, 1, 2, 3, 4`:

Log-likelihood $_{K=1} =$

Log-likelihood $_{K=12} =$

Submit

You have used 0 of 10 attempts

Completing missing entries

0.0/1.0 point (graded)

Now that we have a mixture model, how do we use it to complete a partially observed matrix? We can use the following expression for completing a particular row, say \mathbf{x}_C where the observed values are $\mathbf{i} \in$

In `em.py` implement the function `fill_matrix.py` that takes as input an incomplete matrix `X` and a mixture model, and outputs a completed version of the matrix `X_pred`.

Available Functions: You have access to the NumPy python library as `np`, to the `GaussianMixture` class from `sklearn.mixture` as `GaussianMixture`, and to typing annotation `typing.Tuple` as `Tuple`. You also have access to `scipy.special.logsumexp`.

```
1 def log_gaussian(x: np.ndarray, mean: np.ndarray, var: float) -> float:
2     """Computes the log probability of vector x under a normal distribution
3     Args:
4         x: (d, ) array holding the vector's coordinates
5         mean: (d, ) mean of the gaussian
6         var: variance of the gaussian
7     Returns:
8         float: the log probability
9     """
10    d = len(x)
11    log_prob = -d / 2.0 * np.log(2 * np.pi * var)
```

[Submit](#)

You have used 3 of 50 attempts

Comparing with gold targets

0.0/1.0 point (graded)

Test the accuracy of your predictions against actual target values by loading the complete

`X_gold = np.loadtxt('netflix_complete.txt')` and measuring the root mean square error between the two matrices using `common.rmse(X_gold, X_pred)`. Use your best mixture model for this tab to generate the results.

[Submit](#)

You have used 0 of 10 attempts

More Challenge: Collaborative Filtering Using Matrix Factorization and Neural Networks

Now that you have solved use the Expectation Maximization (EM) algorithm to perform collaborative filtering, you may also want to try solving the same task using matrix factorization as discussed in the lecture. Here are some steps you could follow in that direction:

1. Implement and test matrix factorization directly. (You will actually find it easier to implement a neural network model that works better.)
2. Reformulate the matrix factorization model as a neural network model where you feed the user and movie IDs as input and predict outcome for each pair.
3. Incorporate additional feature information available about the user/movie into the neural network model.

An Application of the EM algorithm: Black Hole Imaging

You may wonder where else the EM algorithm is used.

The following story will draw a connection between the EM algorithm and the first black hole image uncovered to the world on April 10, 2019. The material is mainly based on the work of Katherine Bouman during her PhD at MIT.

[The Sparse](#)[< Previous](#)

ol

[Next >](#)

-
- I've been through your solutions but I still don't understand why the posterior probability output from estep is
-
- ? STAFF - "Completing Missing Entries" Solution looks different than the recitation video explain
I don't see a correlation between the provided solution and explanation in the 2 recitation videos "Gaussian M
-
- 💬 Completing the missing entries
Is it not more sensible to complete the matrix using the (individual) posterior "membership" probabilities for e
-
- ? Correct results marked as wrong
For "Completing missing entries" I am getting the correct results (my shown output is exactly the same as th
-
- ✓ In 'Completing missing entries' question, all the outputs of my code matches with the correct
point?
In 'Completing missing entries' question, all the outputs of my code matches with the correct output howeve
-
- ✓ Log Likelihood not matching on netflix_incomplete.txt, even though the code passed/ ran clea
Log Likelihood not matching on netflix_incomplete.txt, even though the code passed/ ran clean. The code pa
-
- 💬 Solutions
I've been working on this project for about 40 hours now and have reached my limits. The solutions for the p
-
- ? More Challenge: Implemented Code
Will you share the code for "Collaborative Filtering Using Matrix Factorization and Neural Networks" or at lea
-
- ? [STAFF] About "Completing missing entries"

Legal

[Terms of Service & Honor Code](#)

[Privacy Policy](#)

[Accessibility Policy](#)

[Trademark Policy](#)

[Sitemap](#)

[Cookie Policy](#)

[Do Not Sell My Personal Information](#)

Connect

[Blog](#)

[Contact Us](#)

[Help Center](#)

[Security](#)

[Media Kit](#)

