Problem 1. Assume we have \( \{x_1, \ldots, x_N\} \) with \( x \in \mathbb{R}^d \). We model them as independent random variables with \( x_n \sim N(Wz_n, \sigma^2 I) \). The matrix \( W \in \mathbb{R}^{d \times k} \) is unknown and we model \( z_n \sim N(0, I) \). Our goal is to learn a type II maximum likelihood solution for \( W \). That is, find
\[
W' = \arg \max_W \ln p(x_1, \ldots, x_N|W).
\]
Derive an EM algorithm for doing this where \( z_1, \ldots, z_N \) function as the “hidden data”. As a reminder, this involves calculating posteriors for \( z_n \) given \( W^{\text{(old)}} \), taking expectations of the log “complete-data” likelihood and maximizing \( Q(W, W^{\text{(old)}}) \) with respect to \( W \).

Problem 2. Implement a MAP inference algorithm for the matrix completion problem discussed in class. As a reminder, we have \( u_i \sim N(0, \lambda^{-1} I) \) for \( i = 1, \ldots, N_1 \) and \( u \in \mathbb{R}^d \), and \( v_j \sim N(0, \lambda^{-1} I) \) for \( j = 1, \ldots, N_2 \) and \( v \in \mathbb{R}^d \). We are given an \( N_1 \times N_2 \) matrix \( M \) with many missing values. Given the measurement set \( \Omega = \{(i, j) : M_{ij} \text{ is measured}\} \), for each \( (i, j) \in \Omega \) we have \( M_{ij} \sim N(u_i^T v_j, \sigma^2) \).

Run your algorithm on the user-movie ratings dataset provided on the course website. For your algorithm, set \( \sigma^2 = 0.25 \) and \( \lambda = 10 \) (though feel free to try other values). Do the following for \( d = 10, 20, 30 \): Train the model on the larger training set for 100 iterations. For each user-movie pair in the test set, predict the rating by mapping the relevant dot product to the closest integer from 1 to 5. Plot the RMSE of your predictions on this test set as a function of training iteration. On a separate plot show the log joint likelihood as a function of iteration. For \( d = 10 \) implement the Gibbs sampler for this model that we discussed in class. Use a burn-in of 250 iterations. Then for every 25 samples from iteration 251 to 500, calculate the relevant dot products for the data in the test set \( \text{without} \) rounding to an integer. For each user-movie pair in the test set, take the average of these 10 dot products and then round to the nearest integer from 1 to 5. Calculate the RMSE and plot it as a straight line along with the MAP results. On a separate plot show the log joint likelihood as a function every 10 iterations from 1 to 500. (Tip: If there are sampling issues because of the covariance matrix, do an eigendecomposition to make sure the eigenvalues are real and non-negative.)

Using one of your MAP results, pick three movies and for each movie find the 5 closest movies according to Euclidean distance using their respective locations \( v_i \). List the query movie, the five nearest movies and their distances. A mapping from index to movie is given with the data.

Since we did this in class, there’s no need to re-derive it. However, write the update equations and the sampling distributions and briefly indicate what your code is doing with them.