

[← Back to Week 1](#)[X Lessons](#)[This Course: Recommender Systems: Evaluation and Metrics](#)[Next →](#)

This assignment tests your ability to compute basic recommender system metrics from a matrix of ratings and a matrix of predicted values for those ratings.

First, familiarize yourself with the spreadsheet found here:

Metric SpreadSheet Assignment.xlsx

Rows 3-12 provide the ratings matrix for 10 users by 10 movies. For example, cell D10 shows a 3.5, showing that user 2492 gave the movie Forrest Gump a score of 3.5 stars. Many cells are empty -- these reflect movies for which we do not have ratings by those users. To make things easier for later computation, we've also included the count of movie ratings per user, and per movie.

Rows 17-26 provide a similar matrix, but in this case these are predictions from our recommender algorithm. For example, D24 shows 4.2, which means that user 2492 is predicted to give Forrest Gump 4.2 stars. This means the prediction is 0.7 stars higher than the actual rating. Predictions are all between 0 and 5 stars.

Finally, to help you get going, rows 29-38 show you the absolute errors for each prediction. If you look at the formula at the cell B29, `=if(ISNUMBER(B3),abs(B3-B17),"")`, you'll see that we don't just take the absolute value of the difference, but also make sure there is a blank string in any cell that doesn't have a rating to compute from. This is important because the average functions we use will ignore blanks, but would count in zeroes or other values that could be interpreted as numeric.

Note: In an earlier (buggy) version of this spreadsheet, this was expressed as `=if(B3, ...)`. That works here, but does not work below when computing the squared error, since the truth value of B3 is false for both non-numeric values and for the number zero.

Here are the questions we want you to answer (in the quiz that follows). We suggest you pull together all of the answers first, and then paste them into the quiz at the end.

1. Compute the mean absolute error for each user. For example, user 5261 has a mean absolute error of 0.82 (mostly because s/he really had prediction on the Pirates of the Carribean movie).

(a) which users have the highest and lowest MAE, and what are those MAE values?

2. Also compute the mean absolute error associated with the predictions for each movie. For example, Forrest Gump has an MAE of 0.5.

(a) which movies have the highest and lowest MAE, and what are those MAE values?

3. Compute the overall MAE three ways (three decimal places is enough):

(a) the normal MAE, which averages over all predictions

(b) the by-user MAE, which averages the MAEs of each user

(c) the by-movie MAE, which averages the MAEs of each movie

4. Next, create a matrix with the squared errors (instead of absolute errors), and compute the MSE and RMSE (root-mean-squared-error) for each user. Excel and other spreadsheets have a built-in function SQRT() to compute square roots. Be sure that you are computing the MSE first, then taking the square root (if you take the square root of each error before averaging it, you end up with MAE again). For example, user 5136 has an MSE of 0.7911 and an RMSE of .8894.

(a) which users have the highest and lowest RMSE, and what are those RMSE values?

5. Then, compute the overall MSE and RMSE across all predictions (just the normal way, no need to compute by-user or by-movie). Again, remember to average all the squared errors to compute MSE, then take the square root of the MSE to get RMSE.

6. Next, compute the correlation between ratings and predictions using the built-in CORREL() function. First, do this for each user (e.g., the correlation between ratings and predictions for user 5136 is 0.1698 -- only about 17%, and therefore not very good). Then, apply the same correlation to the full matrices.

(a) report the top and bottom correlations (the bottom ones may be negative) by user (both user ID and correlation value).

(b) report the overall correlation between ratings and predictions.

Note that the correlations do not correspond closely to error measures. The errors are about the size of mistakes; the correlations will be high if the ratings and predictions are mostly in the same relative order (i.e., if the predictions do a better job matching relative preferences).

7. We could proceed to other metrics (e.g., large errors or reversals), but we can see that we don't have enough data to see much here (indeed, there is only three errors of two stars or larger), so we'll stop here.



✓ Complete

