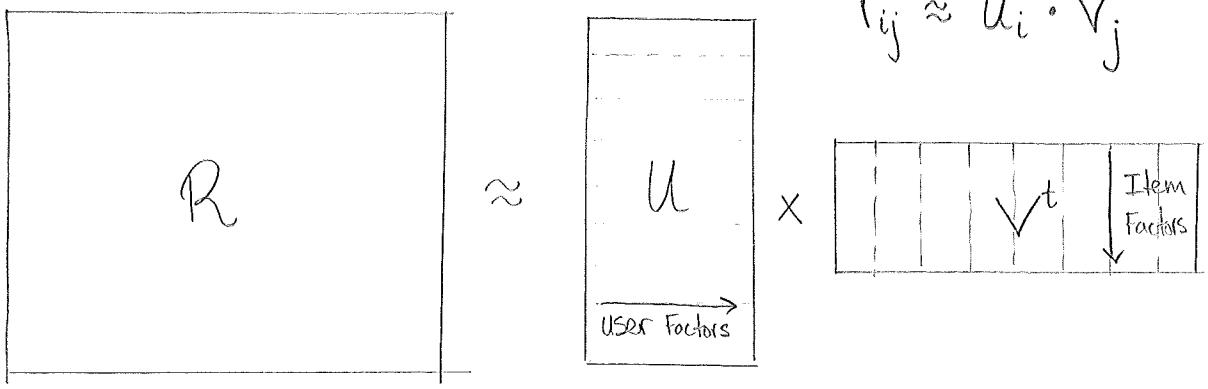
Cross Validation For Matrix Factorization

Reccomender systems involve a peculiar data setup, and this presents some challenges for cross validation.

$$r_{ij} \approx \vec{u}_i \cdot \vec{v}_j$$



- When R consists of explicit ratings, it contains many missing values. Many rows and/or columns may be extremely sparse.

The challenge for model validation comes about if we happen to remove all the ratings in some row or column from the training data.

- If we remove all entries in a row, then we cannot estimate the corresponding user factor.
- If we remove all entries in a column, then we cannot estimate the corresponding item factor.

So, in validating a reccomender we:

- Hold out random ratings.
- Need to ensure that we do not hold out all ratings in a given row or column.