

Dimensions are given below,

But we cannot take all of as there are missing entries.

This is equivalent to,

Similarly,

Dyadic data, each

2 will be absorbed into

If no analytical solution

Gradient descent

1. Start with an initial value of
2. Update by moving along the gradient of the loss function
3. Repeat until converge

As far as I understand, we need at least one data for each user and for each product during training. If we do not have any data for a particular user or a product during training, we cannot predict anything for that corresponding user or product during validation and testing. So while choosing 60% data for training, we must choose all the users and all the products atleast once. If doing so gives us more than 60% data, we split the remaining data equally into test and validation set. However, if it is less, we choose from the remaining data so that the training set becomes 60% and the validation and test sets are 20% each.

Or we could use the global average approach for unseen user with a seen product or an unseen product with a seen user. Problem will occur when we have to predict for an unseen user with an unseen product because even if we have predicted some values for this user with seen products, we can’t learn from them using the global average approach because it is the validation or test set, not the training set.