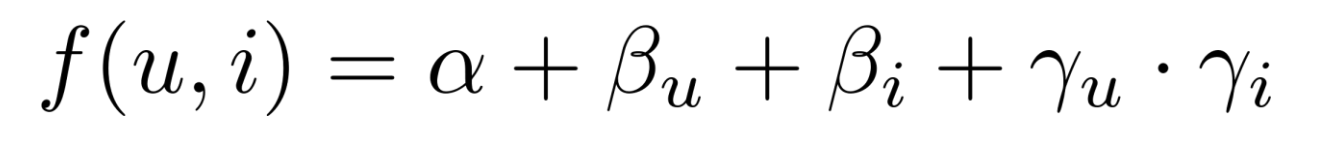
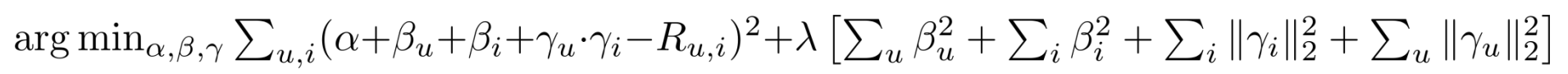
The predictive task remains to be predicting the rating of movie based on available uerID, movieID and rating data. Here based on the linear model, we choose to use the latent factor model with the hope to acquire more accurate prediction. This model was cover by the lecture and in addition to the linear model, this model takes into account the user’s preference and the item’s property. So there are two vectors describing user and item respectively, and we hope the inner product of these two vectors predict the compatibility of the user and item, and thus could perform better than the model which only contain the bias terms.

The model can be written as



and our goal is to minimize the following expression in order to minimize the MSE and to regular the parameters.



The problem with this task is that it is not a convex problem and there is no guarantee that we can get global optimal if we try to solve it with gradient descent. The way to handle it is to solve the linear part first and try to get a good solution for alpha, beta\_u, and beta\_i, and then randomly initialize gamma.\_u and gamma\_i, rather than randomly initialize all parameters.

Then we can user gradient descent and at least we can obtain an suboptimal but still good model to work reasonably well.

Similar to what we did when predict the rating with the linear model, we split the data into training set, validation set, and test set. After the training process, we still need to adjust the value of lamda in the validation process to choose a lamda for the testing process. Continuing our experiment with linear model, and use the alpha, beta\_i and beta\_u from the last experiment and start the gradient descent with gamma\_u and gamma\_i initialized to 0. After the training process, we applied the model to the validation set with lamda ranging from 1 to 10. The value of lamda and the MSE are shown in the graph below.

We can observe from the result of the validation process that the best lamda to use for the testing process is 2 as its corresponding MSE is the lowest.

After calculating the MSE with derived alpha, beta\_u, beta\_i, gamma\_u, gamma\_i with the f(u, i) model mentioned above, we got the MSE on testing set as 0.8134897889764553

**Result**

The result we got from this experiment shows that the latent factor model didn’t outperform the linear model, but their predictions are very close to each other. We think the problem should be in the gradient descent process. Maybe we got out of one local optimal and then fell into another local optimal. Or maybe our gradient descent implementation is not correct as we are not relaying on any third party library in this task. The prediction based on this model didn’t deviate from the linear model too much but takes more effort to implement, This motivates us to explore other method to predict rating based on similarity of users and items.