Homework 6

Ke Wang
kwang337
903058889

2.1 Equation Derivation

$$U_{u,k} = U_{u,k} - \mu \frac{\partial E(U,V)}{\partial U_{u,k}} = U_{u,k} + 2\mu \sum_{(u,i)\epsilon O} (\left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) V_{i,k})$$

$$V_{i,k} = V_{i,k} - \mu \frac{\partial E(U,V)}{\partial V_{i,k}} = V_{i,k} + 2\mu \sum_{(u,i)\epsilon O} (\left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) U_{u,k})$$

2.2 Equation Derivation

$$U_{u,k} = (1 - 2\mu\lambda) U_{u,k} + 2\mu \sum_{(u,i)\epsilon O} (\left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) V_{i,k})$$

$$V_{i,k} = (1 - 2\mu\lambda) V_{i,k} + 2\mu \sum_{(u,i)\epsilon O} (\left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) U_{u,k})$$

PS: I used a slightly different equation in the program, for example, the coefficient 2 is not included (by using a slightly different target function).

2.3.2 Report

r = 1

RMSE	λ = 0.05	λ = 0.1	λ = 0.5
μ = 0.0001	0.9102	0.9157	1.0367
μ = 0.0005	0.9116	0.9203	1.0430
μ = 0.001	0.9122	0.9160	1.0382

r = 3

RMSE	λ = 0.05	λ = 0.1	λ = 0.5
μ = 0.0001	0.8991	0.8902	1.0369
μ = 0.0005	0.8992	0.8896	1.0375
μ = 0.001	0.8969	0.8903	1.0377

r = 5

RMSE	λ = 0.05	λ = 0.1	λ = 0.5
μ = 0.0001	0.8973	0.8855	1.0365
μ = 0.0005	0.8922	0.8879	1.0434
μ = 0.001	0.8957	0.8843	1.0378

Analysis:
1. What do you observe when you vary r? Why?

Generally, given a μ and a λ, a larger r can give better result (smaller RMSE). This is because r represents the number of factors (attributes), and in fact, they are latent factors. Intuitively, a movie has several factors, and user rating can be regarded as his preference (weight on the factors). If we use more factors, then we have a larger chance to distinguish two movies, so we can get a better result.
However, from the result, it seems not true when the regularization factor is set to 0.5, when all the RMSE results are similar. I think this is because the regularization factor is too large so the parameters are too small, so they lead to worse result.

2. Which model is the best? Please describe the best model in Table 4 and explain your choice.

μ	0.001
λ	0.1
r	5
RMSE	0.8843

Table 4: The Best Model

Intuition leads to the choice that a larger r may give a better result. Although we may meet diminishing return when r is large enough, I think in this specific experiment, r=5 is the best. μ is the learning rate. It doesn't matter much when it is small enough to make the algorithm work. From the experiment, it shows that the all three values are fine. So I choose μ to be the largest one to make the program faster.
λ is the regularization factor. It needs to be large enough to avoid over-fitting, but once it is too large, it has a negative impact. From the experiment, $\lambda = 0.05$ and $\lambda = 0.1$ gives good result, and $\lambda = 0.1$ gives a slightly better result, so I choose $\lambda = 0.1$.

3. Suppose you are using regularized MF in real systems, how will you choose parameters? Why?

There are two kinds of parameters in the system, and we should regard them separately. The first group is the coefficient for equations and number of factors. They are fixed once we run the program. First, pick up a small μ by experience. This learning rate doesn't matter very much as long as it is small enough, so we can just pick up one by referring to a typical number. Then we need to decide r and λ. As they will affect each other, we should decide by iteration. We need a large λ to avoid over-fitting, but we don't want the result value to be too small; we desire larger r to get better result, but we may meet diminishing return when r is large enough. r can be chosen by fixing a reasonable λ. We try several possibilities of r, and make a tradeoff between better result and less time.

The other group of parameters is what we are to compute, matrix U and matrix V. We fix the first group of parameters to choose this group. And we should choose good initial values to reduce computation time. The initial value should be determined by real-world constraints (for example, one rating for a movie ranges from 1 to 5) and the parameter from first group (for example, the number of factors).