CSE-6240 HW6 Report

Revant Kumar - 903024989

March 2015

1 Section 2.1

We know that:

$$
\begin{aligned}
E(U,V) &= \sum_{(u,i)\in M} (M_{u,i} - U_u^T V_i)^2 \\
&= \sum_{(u,i)\in M} (M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k})^2
\end{aligned}
$$

Now, we will calculate the partial derivatives as follows:

$\frac{\partial E(U,V)}{\partial U_{u,k}}$ is given by:

$$
\begin{aligned}
\frac{\partial E(U,V)}{\partial U_{u,k}} &= \frac{\partial}{\partial U_{u,k}} \left(\sum_{(u,i)\in M} (M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k})^2 \right) \\
&= \sum_{(u,i)\in M} \frac{\partial}{\partial U_{u,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right)^2 \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \frac{\partial}{\partial U_{u,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \left(\frac{\partial M_{u,i}}{\partial U_{u,k}} - \frac{\partial}{\partial U_{u,k}} \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) (0 - V_{i,k}) \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) V_{i,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k}
\end{aligned}
$$

Now, in the matrix form this can be written as:

$$\frac{\partial E(U,V)}{\partial U} = -2(\mathbf{M} - \mathbf{U}\mathbf{V}^T)\mathbf{V}$$

$\frac{\partial E(U,V)}{\partial V_{j,k}}$ is given by:

$$\begin{aligned}
\frac{\partial E(U,V)}{\partial V_{j,k}} &= \frac{\partial}{\partial V_{j,k}} \left(\sum_{(u,i)\in M} (M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k})^2 \right) \\
&= \sum_{(u,i)\in M} \frac{\partial}{\partial V_{j,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right)^2 \\
&= \sum_{(u:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) \frac{\partial}{\partial V_{j,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) \left(\frac{\partial M_{u,i}}{\partial V_{j,k}} - \frac{\partial}{\partial V_{j,k}} \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) (0 - U_{u,k}) \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k}V_{i,k} \right) U_{u,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) U_{u,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) U_{u,k}
\end{aligned}$$

Now, in the matrix form this can be written as:

$$\frac{\partial E(U,V)}{\partial V} = -2(\mathbf{M} - \mathbf{U}\mathbf{V}^T)^T\mathbf{U}$$

So, now that we have found the gradients, we will update $U_{u,k}\&V_{j,k}$ as follows:

$$\begin{aligned}
U_{u,k} = U_{u,k} - \mu \frac{\partial E(U,V)}{\partial U_{u,k}} \\
&= U_{u,k} - \mu \left(-2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k} \right) \\
&= U_{u,k} + 2\mu \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k}
\end{aligned}$$

$$
\begin{aligned}
V_{j,k} &= V_{j,k} - \mu \frac{\partial E(U,V)}{\partial V_{j,k}} \\
&= V_{j,k} - \mu \left(-2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) U_{u,k} \right) \\
&= V_{j,k} + 2\mu \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) U_{u,k}
\end{aligned}
$$

2 Section 2.2

Now, we will calculate the partial derivatives as follows:

$\frac{\partial E(U,V)}{\partial U_{u,k}}$ is given by:

$$
\begin{aligned}
\frac{\partial E(U,V)}{\partial U_{u,k}} &= \frac{\partial}{\partial U_{u,k}} \left(\sum_{(u,i)\in M} (M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k})^2 + \lambda \sum_{u,k} U_{u,k}^2 + \lambda \sum_{i,k} V_{i,k}^2 \right) \\
&= \frac{\partial}{\partial U_{u,k}} \left(\sum_{(u,i)\in M} (M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k})^2 \right) + \frac{\partial}{\partial U_{u,k}} \left(\lambda \sum_{u,k} U_{u,k}^2 \right) + \frac{\partial}{\partial U_{u,k}} \left(\lambda \sum_{i,k} V_{i,k}^2 \right) \\
&= \sum_{(u,i)\in M} \frac{\partial}{\partial U_{u,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right)^2 + \lambda \sum_{u,k} \frac{\partial}{\partial U_{u,k}} \left(U_{u,k}^2 \right) + \lambda \sum_{i,k} \frac{\partial}{\partial U_{u,k}} \left(V_{i,k}^2 \right) \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \frac{\partial}{\partial U_{u,k}} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) + \lambda (2U_{u,k}) + 0 \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) \left(\frac{\partial M_{u,i}}{\partial U_{u,k}} - \frac{\partial}{\partial U_{u,k}} \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) + 2\lambda U_{u,k} \\
&= \sum_{(i:M(u,i)>0)} 2 \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) (0 - V_{i,k}) + 2\lambda U_{u,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - \sum_{k=1}^{r} U_{u,k} V_{i,k} \right) V_{i,k} + 2\lambda U_{u,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k} + 2\lambda U_{u,k} \\
&= -2 \sum_{(i:M(u,i)>0)} \left(M_{u,i} - U_u^T V_i \right) V_{i,k} + 2\lambda U_{u,k}
\end{aligned}
$$

Now, in the matrix form this can be written as:

$$
\frac{\partial E(U,V)}{\partial U} = -2(\mathbf{M} - \mathbf{U}\mathbf{V}^T)\mathbf{V} + 2\lambda \mathbf{U}
$$

$\frac{\partial E(U,V)}{\partial V_{j,k}}$ is given by:

$$
\begin{aligned}
\frac{\partial E(U,V)}{\partial V_{j,k}} &= \frac{\partial}{\partial V_{j,k}}\left(\sum_{(u,i)\in M}(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k})^2+\lambda\sum_{u,k}U_{u,k}^2+\lambda\sum_{i,k}V_{i,k}^2\right)\\
&= \frac{\partial}{\partial V_{j,k}}\left(\sum_{(u,i)\in M}(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k})^2\right)+\frac{\partial}{\partial V_{j,k}}\left(\lambda\sum_{u,k}U_{u,k}^2\right)+\frac{\partial}{\partial V_{j,k}}\left(\lambda\sum_{i,k}V_{i,k}^2\right)\\
&= \sum_{(u,i)\in M}\frac{\partial}{\partial V_{j,k}}\left(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k}\right)^2+\lambda\sum_{u,k}\frac{\partial}{\partial V_{j,k}}\left(U_{u,k}^2\right)+\lambda\sum_{i,k}\frac{\partial}{\partial V_{j,k}}\left(V_{i,k}^2\right)\\
&= \sum_{(i:M(u,i)>0)}2\left(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k}\right)\left(\frac{\partial M_{u,i}}{\partial V_{j,k}}-\frac{\partial}{\partial V_{j,k}}\sum_{k=1}^{r}U_{u,k}V_{i,k}\right)+0+\lambda(2V_{j,k})\\
&= \sum_{(i:M(u,i)>0)}2\left(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k}\right)(0-U_{u,k})+2\lambda V_{j,k}\\
&= -2\sum_{(i:M(u,i)>0)}\left(M_{u,i}-\sum_{k=1}^{r}U_{u,k}V_{i,k}\right)U_{u,k}+2\lambda V_{j,k}\\
&= -2\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)U_{u,k}+2\lambda V_{j,k}\\
&= -2\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)U_{u,k}+2\lambda V_{j,k}
\end{aligned}
$$

Now, in the matrix form this can be written as:

$$
\frac{\partial E(U,V)}{\partial V} = -2(\mathbf{M}-\mathbf{U}\mathbf{V}^T)^T\mathbf{U}+2\lambda\mathbf{V}
$$

So, now that we have found the gradients, we will update $U_{u,k}\&V_{j,k}$ as follows:

$$
\begin{aligned}
U_{u,k} &= U_{u,k}-\mu\frac{\partial E(U,V)}{\partial U_{u,k}}\\
&= U_{u,k}-\mu\left(-2\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)V_{i,k}+2\lambda U_{u,k}\right)\\
&= (1-2\lambda)U_{u,k}+2\mu\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)V_{i,k}
\end{aligned}
$$

$$
\begin{aligned}
V_{j,k} &= V_{j,k}-\mu\frac{\partial E(U,V)}{\partial V_{j,k}}\\
&= V_{j,k}-\mu\left(-2\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)U_{u,k}+2\lambda V_{j,k}\right)\\
&= (1-2\lambda)V_{j,k}+2\mu\sum_{(i:M(u,i)>0)}\left(M_{u,i}-U_u^TV_i\right)U_{u,k}
\end{aligned}
$$

3 Section 2.3.2

RMSE $r = 1$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = 0.0001$	0.9187	0.9184	0.9199
$\mu = 0.0005$	0.9967	0.9958	0.9841
$\mu = 0.001$	3.7921	3.9826	4.0196

RMSE $r = 3$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = 0.0001$	0.9167	0.9183	0.9176
$\mu = 0.0005$	0.9334	0.9346	0.9190
$\mu = 0.001$	4.5996	4.6174	4.5064

RMSE $r = 5$	$\lambda = 0.05$	$\lambda = 0.1$	$\lambda = 0.5$
$\mu = 0.0001$	0.9202	0.9211	0.9210
$\mu = 0.0005$	0.9485	0.9430	0.9295
$\mu = 0.001$	3.1634	3.2539	3.0749

Now, answers to the questions in the assignment:

1. When we vary r, we observe the following:
   - As we increase the value of r from 1 to 3, we observe that the RMSE decreases. However, when we increase the value of r from 3 to 5, we observe that the RMSE increases.
   - The time taken to run the program increases as we increase the value of r. This is due to the increased dimensions of the user matrix and movies matrix.
   - As we increase r, the best RMSEs are obtained when μ is 0.0001 and the worst RMSEs when μ is 0.001. In case of μ being 0.001, the learning step is too big, and as a result gradient descent is not able to converge to the global minima.
   - Also, in general, we will observe that by increasing the value of r above 5, the value of RMSE increases for a fixed value of μ, λ and number of iterations i.e. the stopping criteria.

2. Best Model is given by the following table:

Best Model	Values
μ	0.0001
λ	0.05
r	3
RMSE	0.9167

   I choose this model as the best model because it gave the minimum average RMSE after preforming cross validation for various values of r, μ and λ.

3. In real systems, when we are using regularized Matrix Factorization, we will use the technique of cross-validation to choose parameters. We will use the 10-fold cross validation on the given training set for the various values of r, μ and λ. The parameters which will give the minimum RMSE will be chosen for the final model to predict the new ratings. Taking the parameters corresponding to the least RMSE will ensure that our final model is more accurate in predicting new ratings and thus produces less error. Also, this will choose the model which will not over-fit the given data.