1. Matrix factorization

We want to find minimum of the following function:

$$RMSE(U,M) = \frac{1}{2}\sum_{i=1}^{N_u}\left(R_{i,:} - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right) \cdot \left(R_{i,:} - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T$$

$$h(x) = \begin{cases} 1, & x > 0 \\ 0, & x \le 0 \end{cases}$$

$D(x)$ is a diagonal matrix where the elements of the vector x are placed in the main diagonal. Let's consider M as a constant and let's represent $RMSE$ as a function of U:

$$RMSE(U) = \frac{1}{2}\sum_{i=1}^{N_u}\left(A - {U_{:,i}}^T \cdot B\right) \cdot \left(A - {U_{:,i}}^T \cdot B\right)^T$$

$$A = R_{i,:}$$

$$B = M \cdot D\left(h\left(R_{i,:}\right)\right)$$

Let's find derivative of $RMSE(U_{:,i})$:

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = \frac{d}{dU_{:,i}}\left(\frac{1}{2} \cdot \left(A - {U_{:,i}}^T \cdot B\right) \cdot \left(A - {U_{:,i}}^T \cdot B\right)^T\right)$$

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = \frac{1}{2} \cdot \frac{d}{dU_{:,i}}\left(A \cdot A^T - A \cdot \left({U_{:,i}}^T \cdot B\right)^T - {U_{:,i}}^T \cdot B \cdot A^T + {U_{:,i}}^T \cdot B \cdot \left({U_{:,i}}^T \cdot B\right)^T\right)$$

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = \frac{1}{2} \cdot \frac{d}{dU_{:,i}}\left(A \cdot A^T - A \cdot B^T \cdot U_{:,i} - {U_{:,i}}^T \cdot B \cdot A^T + {U_{:,i}}^T \cdot B \cdot B^T \cdot U_{:,i}\right)$$

We will use the following identities:

$$\frac{d}{dx}\left(x^T \cdot a\right) = \frac{d}{dx}\left(a^T \cdot x\right) = a$$

$$\frac{d}{dx}\left(x^T \cdot B \cdot x\right) = \left(B + B^T\right) \cdot x$$

Let's continue with the derivation of $RMSE(U_{:,i})$:

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = \frac{1}{2} \cdot \left(0 - (A \cdot B^T)^T - B \cdot A^T + (B \cdot B^T + (B \cdot B^T)^T) \cdot U_{:,i}\right)$$

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = \frac{1}{2} \cdot \left(-2 \cdot B \cdot A^T + 2 \cdot B \cdot B^T \cdot U_{:,i}\right)$$

$$\frac{d}{dU_{:,i}}RMSE(U_{:,i}) = B \cdot B^T \cdot U_{:,i} - B \cdot A^T$$

Let's replace A and B with the original expressions:

$$\frac{d}{dU_{:,i}} RMSE(U_{:,i}) = M \cdot D\left(h(R_{i,:})\right) \cdot \left(M \cdot D\left(h(R_{i,:})\right)\right)^T \cdot U_{:,i} - M \cdot D\left(h(R_{i,:})\right) \cdot {R_{i,:}}^T$$

$$\frac{d}{dU_{:,i}} RMSE(U_{:,i}) = M \cdot D\left(h(R_{i,:})\right) \cdot M^T \cdot U_{:,i} - M \cdot D\left(h(R_{i,:})\right) \cdot {R_{i,:}}^T$$

This expression can be used in Gradient Descent. However, we can even directly find the optimal solution:

$$M \cdot D\left(h(R_{i,:})\right) \cdot M^T \cdot U_{:,i} - M \cdot D\left(h(R_{i,:})\right) \cdot {R_{i,:}}^T = 0$$

$$U_{:,i} = \left(M \cdot D\left(h(R_{i,:})\right) \cdot M^T\right)^{-1} \cdot M \cdot D\left(h(R_{i,:})\right) \cdot {R_{i,:}}^T$$

In a similar way we can fix $U_{:,i}$ and we can obtain an expression of $RMSE$ as a function of $M_{:,i}$. We should begin with the following expression:

$$RMSE(U, M) = \frac{1}{2} \sum_{j=1}^{N_m} \left({R_{:,j}}^T - {M_{:,j}}^T \cdot U \cdot D\left(h(R_{:,j})\right)\right) \cdot \left({R_{:,j}}^T - {M_{:,j}}^T \cdot U \cdot D\left(h(R_{:,j})\right)\right)^T$$

$$A = {R_{:,j}}^T$$

$$B = U \cdot D\left(h(R_{:,j})\right)$$

Let's find derivative of $RMSE(M_{:,j})$:

$$\frac{d}{dM_{:,j}} RMSE(M_{:,j}) = \frac{d}{dU_{:,i}} \left(\frac{1}{2} \cdot (A - {M_{:,j}}^T \cdot B) \cdot (A - {M_{:,j}}^T \cdot B)^T\right)$$

$$\frac{d}{dU_{:,i}} RMSE(M_{:,j}) = B \cdot B^T \cdot M_{:,j} - B \cdot A^T$$

$$\frac{d}{dM_{:,j}} RMSE(M_{:,j}) = U \cdot D\left(h(R_{:,j})\right) \cdot U^T \cdot M_{:,j} - U \cdot D\left(h(R_{:,j})\right) \cdot R_{:,j}$$

The optimal solution is:

$$M_{:,j} = \left(U \cdot D\left(h(R_{:,j})\right) \cdot U^T\right)^{-1} \cdot U \cdot D\left(h(R_{:,j})\right) \cdot R_{:,j}$$

2. Matrix factorization with penalization term

We want to find minimum of the following function:

$$RMSE^{(2)}(U,M) = \frac{1}{2}\sum_{i=1}^{N_u}\left(R_{i,:} - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)\cdot\left(R_{i,:} - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T + \frac{\lambda_u}{2} \cdot \sum_{i=1}^{N_u} nru_i \cdot U_{:,i}{}^T \cdot U_{:,i} + \frac{\lambda_m}{2} \cdot \sum_{j=1}^{N_m} nrm_j \cdot M_{:,j}{}^T \cdot M_{:,j}$$

$$RMSE^{(2)}(U,M) = RMSE(U,M) + \frac{\lambda_u}{2} \cdot \sum_{i=1}^{N_u} nru_i \cdot U_{:,i}{}^T \cdot U_{:,i} + \frac{\lambda_m}{2} \cdot \sum_{j=1}^{N_m} nrm_j \cdot M_{:,j}{}^T \cdot M_{:,j}$$

Let's find the derivative w.r.t. $U_{:,i}$:

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = \frac{d}{dU_{:,i}} RMSE(U_{:,i}) + \frac{\lambda_u}{2} \cdot \frac{d}{dU_{:,i}}\left(nru_i \cdot U_{:,i}{}^T \cdot U_{:,i}\right) + \frac{\lambda_m}{2} \cdot \frac{d}{dU_{:,i}}\left(\sum_{j=1}^{N_m} nrm_j \cdot M_{:,j}{}^T \cdot M_{:,j}\right)$$

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = \frac{d}{dU_{:,i}} RMSE(U_{:,i}) + \frac{\lambda_u \cdot nru_i}{2} \cdot 2 \cdot U_{:,i} + 0$$

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = \frac{d}{dU_{:,i}} RMSE(U_{:,i}) + \lambda_u \cdot nru_i \cdot U_{:,i}$$

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot M^T \cdot U_{:,i} - M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot R_{i,:}{}^T + \lambda_u \cdot nru_i \cdot U_{:,i}$$

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = \left(M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot M^T + \lambda_u \cdot nru_i \cdot D\left(I(nlf)\right)\right) \cdot U_{:,i} - M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot R_{i,:}{}^T$$

where nlf is the number of latent factors used in the factorization. Let's find the optimal solution:

$$\frac{d}{dU_{:,i}} RMSE^{(2)}(U_{:,i}) = 0$$

$$U_{:,i} = \left(M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot M^T + \lambda_u \cdot nru_i \cdot D\left(I(nlf)\right)\right)^{-1} \cdot M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot R_{i,:}{}^T$$

Let's find the derivative w.r.t. $M_{:,j}$:

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j})$$
$$= \frac{d}{dM_{:,j}} RMSE(M_{:,j}) + \frac{\lambda_u}{2} \cdot \frac{d}{dM_{:,j}} \left(\sum_{i=1}^{N_u} nru_i \cdot U_{:,i}{}^T \cdot U_{:,i} \right) + \frac{\lambda_m}{2} \cdot \frac{d}{dM_{:,j}} \left(nrm_j \cdot M_{:,j}{}^T \cdot M_{:,j} \right)$$

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j}) = \frac{d}{dM_{:,j}} RMSE(M_{:,j}) + 0 + \frac{\lambda_m \cdot nrm_j}{2} \cdot 2 \cdot M_{:,j}$$

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j}) = \frac{d}{dM_{:,j}} RMSE(M_{:,j}) + \lambda_m \cdot nrm_j \cdot M_{:,j}$$

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j}) = U \cdot D\left(h(R_{:,j})\right) \cdot U^T \cdot M_{:,j} - U \cdot D\left(h(R_{:,j})\right) \cdot R_{:,j} + \lambda_m \cdot nrm_j \cdot M_{:,j}$$

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j}) = \left(U \cdot D\left(h(R_{:,j})\right) \cdot U^T + \lambda_m \cdot nrm_j \cdot D\left(I(nlf)\right) \right) \cdot M_{:,j} - U \cdot D\left(h(R_{:,j})\right) \cdot R_{:,j}$$

Let’s find the optimal solution:

$$\frac{d}{dM_{:,j}} RMSE^{(2)}(M_{:,j}) = 0$$

$$M_{:,j} = \left(U \cdot D\left(h(R_{:,j})\right) \cdot U^T + \lambda_m \cdot nrm_j \cdot D\left(I(nlf)\right) \right)^{-1} \cdot U \cdot D\left(h(R_{:,j})\right) \cdot R_{:,j}$$

3. Matrix factorization with bias and penalization term

We want to find minimum of the following function:

$$RMSE^{(3)}(U,M) = \frac{1}{2}\sum_{i=1}^{N_u}\left(R_{i,:} - \mu^T - {bu_i}^T - bm^T - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)$$
$$\cdot\left(R_{i,:} - \mu^T - {bu_i}^T - bm^T - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T + \frac{\lambda_u}{2}\cdot\sum_{i=1}^{N_u} nru_i \cdot {U_{:,i}}^T \cdot U_{:,i} + \frac{\lambda_m}{2}$$
$$\cdot\sum_{j=1}^{N_m} nrm_j \cdot {M_{:,j}}^T \cdot M_{:,j}$$

where μ is a column vector of length N_m in which all values are equal and represent the global bias, bu_i is a column vector of length N_m in which all values are equal and represent the bias of the u-th user and bm is a column vector of length N_m in which the values represent the bias of each item. Let's represent $RMSE^{(3)}$ as a function of bm and let's simplify this equation in the following way:

$$RMSE^{(3)}(bm) = \frac{1}{2}\sum_{i=1}^{N_u}(A - bm^T)\cdot(A - bm^T)^T + \frac{\lambda_u}{2}\cdot\sum_{i=1}^{N_u} nru_i \cdot {U_{:,i}}^T \cdot U_{:,i} + \frac{\lambda_m}{2}\cdot\sum_{j=1}^{N_m} nrm_j \cdot {M_{:,j}}^T \cdot M_{:,j}$$

$$A = R_{i,:} - \mu^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)$$

Let's find the derivative w.r.t. bm:

$$\frac{d}{dbm}RMSE^{(3)}(bm) = \frac{d}{dbm}\left(\frac{1}{2}\sum_{i=1}^{N_u}(A - bm^T)\cdot(A - bm^T)^T\right) + 0 + 0$$

$$\frac{d}{dbm}RMSE^{(3)}(bm) = \frac{1}{2}\cdot\frac{d}{dbm}\left(\sum_{i=1}^{N_u} A\cdot A^T - A\cdot bm - bm^T\cdot A^T + bm^T\cdot bm\right)$$

$$\frac{d}{dbm}RMSE^{(3)}(bm) = \frac{1}{2}\cdot\left(\sum_{i=1}^{N_u} 0 - A^T - A^T + 2\cdot bm\right) = \sum_{i=1}^{N_u} bm - A^T = N_u\cdot bm - \sum_{i=1}^{N_u} A^T$$

$$\frac{d}{dbm}RMSE^{(3)}(bm) = N_u\cdot bm - \sum_{i=1}^{N_u}\left(R_{i,:} - \mu^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T$$

The optimal solution is:

$$bm = \frac{1}{N_u}\cdot\sum_{i=1}^{N_u}\left(R_{i,:} - \mu^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T$$

In a similar way we can find the derivative w.r.t. bu

$$\frac{d}{dbu}RMSE^{(3)}(bu) = N_m \cdot bu - \sum_{j=1}^{N_m}\left(R_{:,j}{}^T - \mu^T - bm_j{}^T - M_{:,j}{}^T \cdot U \cdot D\left(h\left(R_{:,j}\right)\right)\right)^T$$

The optimal solution is:

$$bu = \frac{1}{N_m} \cdot \sum_{j=1}^{N_m}\left(R_{:,j}{}^T - \mu^T - bm_j{}^T - M_{:,j}{}^T \cdot U \cdot D\left(h\left(R_{:,j}\right)\right)\right)^T$$

The derivative w.r.t. μ is:

$$\frac{d}{d\mu}RMSE^{(3)}(\mu) = N_u \cdot \mu - \sum_{i=1}^{N_u}\left(R_{i,:} - bm^T - bu_i{}^T - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T$$

The optimal solution is:

$$\mu = \frac{1}{N_u} \cdot \sum_{i=1}^{N_u}\left(R_{i,:} - bm^T - bu_i{}^T - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T$$

The updated derivative w.r.t. $U_{:,i}$ is:

$$\frac{d}{dU_{:,i}}RMSE^{(3)}(U_{:,i}) = \left(M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot M^T + \lambda_u \cdot nru_i \cdot D\left(I(nlf)\right)\right) \cdot U_{:,i} - M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot \left(R_{i,:}{}^T - \mu - bu_i - bm\right)$$

The optimal solution for $U_{:,i}$ is:

$$U_{:,i} = \left(M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot M^T + \lambda_u \cdot nru_i \cdot D\left(I(nlf)\right)\right)^{-1} \cdot M \cdot D\left(h\left(R_{i,:}\right)\right) \cdot \left(R_{i,:}{}^T - \mu - bu_i - bm\right)$$

The updated derivative w.r.t. $M_{:,j}$ is:

$$\frac{d}{dM_{:,j}}RMSE^{(3)}(M_{:,j}) = \left(U \cdot D\left(h\left(R_{:,j}\right)\right) \cdot U^T + \lambda_m \cdot nrm_j \cdot D\left(I(nlf)\right)\right) \cdot M_{:,j} - U \cdot D\left(h\left(R_{:,j}\right)\right) \cdot \left(R_{:,j} - \mu - bu - bm_j\right)$$

The optimal solution for $M_{:,j}$ is:

$$M_{:,j} = \left(U \cdot D\left(h\left(R_{:,j}\right)\right) \cdot U^T + \lambda_m \cdot nrm_j \cdot D\left(I(nlf)\right)\right)^{-1} \cdot U \cdot D\left(h\left(R_{:,j}\right)\right) \cdot \left(R_{:,j} - \mu - bu - bm_j\right)$$

4. Matrix factorization with bias, penalization term and friendship term

We want to find minimum of the following function:

$$RMSE^{(4)}(U,M) = \frac{1}{2}\sum_{i=1}^{N_u}\left(R_{i,:} - \mu^T - bu_i{}^T - bm^T - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right) \cdot \left(R_{i,:} - \mu^T - bu_i{}^T - bm^T - U_{:,i}{}^T \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)^T + \frac{\lambda_u}{2} \cdot \sum_{i=1}^{N_u} nru_i \cdot U_{:,i}{}^T \cdot U_{:,i} + \frac{\lambda_m}{2} \cdot \sum_{j=1}^{N_m} nrm_j \cdot M_{:,j}{}^T \cdot M_{:,j} + \frac{\lambda_t}{2} \cdot \sum_{i=1}^{N_u}\left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)^T \cdot \left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)$$

$$RMSE^{(4)}(U,M) = RMSE^{(2)}(U,M) + \frac{\lambda_t}{2} \cdot \sum_{i=1}^{N_u}\left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)^T \cdot \left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)$$

where nf_i is the number of friends of the i-th user, $F_{i,j} = 1$ if i-th and j-th users are friends and 0 otherwise, $I(n)$ is a column matrix with n ones inside. Let's find the derivative w.r.t. $U_{:,i}$:

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \frac{\lambda_t}{2} \cdot \frac{d}{dU_{:.i}} \sum_{i=1}^{N_u}\left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)^T \cdot \left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)\right)$$

Let's reformulate this expression:

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \frac{\lambda_t}{2} \cdot \frac{d}{dU_{:.i}} \sum_{i=1}^{N_u}\left(U_{:,i} - A\right)^T \cdot \left(U_{:,i} - A\right)$$

$$A = \frac{1}{nf_i} \cdot U \cdot D\left(F_{i,:}\right) \cdot I(N_u)$$

We can include U in the constant expression because the expression does not depend on $U_{:.i}$ (a user cannot be a friend with himself).

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \frac{\lambda_t}{2} \cdot \frac{d}{dU_{:.i}} \sum_{i=1}^{N_u}\left(U_{:,i}{}^T \cdot U_{:,i} - U_{:,i}{}^T \cdot A - A^T \cdot U_{:,i}\right)$$

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \frac{\lambda_t}{2} \cdot (2 \cdot U_{:,i} - 2 \cdot A)$$

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \lambda_t \cdot (U_{:,i} - A)$$

$$\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) = \frac{d}{dU_{:.i}} RMSE^{(3)}(U_{:.i}) + \lambda_t \cdot \left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D(F_{i,:}) \cdot I(N_u) \right)$$

$$\begin{aligned} &\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) \\ &= \left(M \cdot D\left(h(R_{i,:}) \right) \cdot M^T + \lambda_u \cdot nru_i \cdot D(I(nlf)) \right) \cdot U_{:,i} - M \cdot D\left(h(R_{i,:}) \right) \\ &\cdot (R_{i,:}{}^T - \mu - bu_i - bm) + \lambda_t \cdot \left(U_{:,i} - \frac{1}{nf_i} \cdot U \cdot D(F_{i,:}) \cdot I(N_u) \right) \end{aligned}$$

$$\begin{aligned} &\frac{d}{dU_{:.i}} RMSE^{(4)}(U_{:.i}) \\ &= \left(M \cdot D\left(h(R_{i,:}) \right) \cdot M^T + (\lambda_u \cdot nru_i + \lambda_t) \cdot D(I(nlf)) \right) \cdot U_{:,i} - M \cdot D\left(h(R_{i,:}) \right) \\ &\cdot (R_{i,:}{}^T - \mu - bu_i - bm) - \frac{\lambda_t}{nf_i} \cdot U \cdot D(F_{i,:}) \cdot I(N_u) \end{aligned}$$

The optimal solution is:

$$\begin{aligned} U_{:.i} = &\left(M \cdot D\left(h(R_{i,:}) \right) \cdot M^T + (\lambda_u \cdot nru_i + \lambda_t) \cdot D(I(nlf)) \right)^{-1} \\ &\cdot \left(M \cdot D\left(h(R_{i,:}) \right) \cdot (R_{i,:}{}^T - \mu - bu_i - bm) - \frac{\lambda_t}{nf_i} \cdot U \cdot D(F_{i,:}) \cdot I(N_u) \right) \end{aligned}$$

5. Matrix factorization with exponential ratings

To fit a functional form:

$$y = A \cdot e^{B \cdot x}$$

we can use the following cost function:

$$\sum_{i=1}^{n} y_i \cdot (\ln y_i - \ln A - B \cdot x_i)^2$$

Let's write this in vector form and adjust it to our problem:

$$RMSE^{(5)}(U, M) = \frac{1}{2} \sum_{i=1}^{N_u} \left(\ln(R_{i,:}) - \mu^T - {bu_i}^T - bm^T - {U_{:,i}}^T \cdot M \cdot D\left(h(R_{i,:})\right) \right) \cdot D(R_{i,:}) \cdot \left(\ln(R_{i,:}) - \mu^T - {bu_i}^T - bm^T - {U_{:,i}}^T \cdot M \cdot D\left(h(R_{i,:})\right) \right)^T$$

We can add the penalization term and the friendship term, but because those expressions can be considered as independent, we don't add complexity to this formula in this phase. Let's simplify the expression:

$$RMSE^{(5)}(U, M) = \frac{1}{2} \sum_{i=1}^{N_u} \left(A - {U_{:,i}}^T \cdot B \right) \cdot C \cdot \left(A - {U_{:,i}}^T \cdot B \right)^T$$

$$A = \ln(R_{i,:}) - \mu^T - {bu_i}^T - bm^T$$

$$B = M \cdot D\left(h(R_{i,:})\right)$$

$$C = D(R_{i,:})$$

Let's find the derivative w.r.t $U_{:,i}$:

$$\frac{d}{dU_{:.i}} RMSE^{(5)}(U_{:.i}) = \frac{1}{2} \frac{d}{dU_{:.i}} \left(A \cdot C \cdot A^T - A \cdot C \cdot B^T \cdot U_{:,i} - {U_{:,i}}^T \cdot B \cdot C \cdot A^T + {U_{:,i}}^T \cdot B \cdot C \cdot B^T \cdot U_{:,i} \right)$$

$$\frac{d}{dU_{:.i}} RMSE^{(5)}(U_{:.i}) = \frac{1}{2} \cdot \left(0 - (A \cdot C \cdot B^T)^T - B \cdot C \cdot A^T + (B \cdot C \cdot B^T + (B \cdot C \cdot B^T)^T) \cdot U_{:,i} \right)$$

$$\frac{d}{dU_{:.i}} RMSE^{(5)}(U_{:.i}) = \frac{1}{2} \cdot \left(-2 \cdot B \cdot C \cdot A^T + 2 \cdot B \cdot C \cdot B^T \cdot U_{:,i} \right)$$

$$\frac{d}{dU_{:.i}} RMSE^{(5)}(U_{:.i}) = B \cdot C \cdot B^T \cdot U_{:,i} - B \cdot C \cdot A^T$$

$$\frac{d}{dU_{:.i}} RMSE^{(5)}(U_{:.i}) = M \cdot D\left(h(R_{i,:})\right) \cdot D(R_{i,:}) \cdot D\left(h(R_{i,:})\right) \cdot M^T \cdot U_{:,i} - M \cdot D\left(h(R_{i,:})\right) \cdot D(R_{i,:}) \cdot \left(\ln(R_{i,:}) - \mu^T - {bu_i}^T - bm^T \right)^T$$

$$D\left(h\left(R_{i,:}\right)\right)\cdot D\left(R_{i,:}\right)=D\left(R_{i,:}\right)\cdot D\left(h\left(R_{i,:}\right)\right)=D\left(R_{i,:}\right)$$

$$\frac{d}{dU_{:,i}}RMSE^{(5)}(U_{:,i})= M\cdot D\left(R_{i,:}\right)\cdot M^T\cdot U_{:,i}-M\cdot D\left(R_{i,:}\right)\cdot\left(\ln\left(R_{i,:}\right)-\mu^T-{bu_i}^T-bm^T\right)^T$$

The optimal solution is:

$$U_{:,i}=\left(M\cdot D\left(R_{i,:}\right)\cdot M^T\right)^{-1}\cdot M\cdot D\left(R_{i,:}\right)\cdot\left(\ln\left(R_{i,:}\right)-\mu^T-{bu_i}^T-bm^T\right)^T$$

In a similar way:

$$\frac{d}{dM_{:,j}}RMSE^{(5)}\left(M_{:,j}\right)= U\cdot D\left(R_{:,j}\right)\cdot U^T\cdot M_{:,j}-U\cdot D\left(R_{:,j}\right)\cdot\left(\ln\left({R_{:,j}}^T\right)-\mu^T-bu^T-{bm_j}^T\right)^T$$

The optimal solution is:

$$M_{:,j}=\left(U\cdot D\left(R_{:,j}\right)\cdot U^T\right)^{-1}\cdot U\cdot D\left(R_{:,j}\right)\cdot\left(\ln\left({R_{:,j}}^T\right)-\mu^T-bu^T-{bm_j}^T\right)^T$$

Let’s find the derivative w.r.t. *bm*:

$$RMSE^{(5)}(bm)=\frac{1}{2}\sum_{i=1}^{N_u}(A-bm^T)\cdot C\cdot(A-bm^T)^T$$

$$A=\ln\left(R_{i,:}\right)-\mu^T-{bu_i}^T-{U_{:,i}}^T\cdot M\cdot D\left(h\left(R_{i,:}\right)\right)$$

$$C=D\left(R_{i,:}\right)$$

$$\frac{d}{dbm}RMSE^{(5)}(bm)=\frac{1}{2}\cdot\frac{d}{dbm}\sum_{i=1}^{N_u}(A-bm^T)\cdot C\cdot(A-bm^T)^T$$

$$\frac{d}{dbm}RMSE^{(5)}(bm)=\frac{1}{2}\cdot\frac{d}{dbm}\sum_{i=1}^{N_u}(A\cdot C\cdot A^T-A\cdot C\cdot bm-bm^T\cdot C\cdot A^T+bm^T\cdot C\cdot bm)$$

$$\frac{d}{dbm}RMSE^{(5)}(bm)=\frac{1}{2}\cdot\sum_{i=1}^{N_u}(0-C^T\cdot A^T-C\cdot A^T+(C+C^T)\cdot bm)$$

$$\frac{d}{dbm}RMSE^{(5)}(bm)=\frac{1}{2}\cdot\sum_{i=1}^{N_u}(-2\cdot C\cdot A^T+2\cdot C\cdot bm)=\sum_{i=1}^{N_u}C\cdot(bm-A^T)$$

$$\frac{d}{dbm}RMSE^{(5)}(bm)=\sum_{i=1}^{N_u}D\left(R_{i,:}\right)\cdot\left(bm-\left(\ln\left(R_{i,:}\right)-\mu^T-{bu_i}^T-{U_{:,i}}^T\cdot M\cdot D\left(h\left(R_{i,:}\right)\right)\right)^T\right)$$

The optimal solution is given by:

$$bm = \left(\sum_{i=1}^{N_u} D(R_{i,:})\right)^{-1} \cdot \sum_{i=1}^{N_u} D(R_{i,:}) \cdot \left(\ln(R_{i,:}) - \mu^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h(R_{i,:})\right)\right)^T$$

The derivative w.r.t. bu is:

$$\frac{d}{dbu} RMSE^{(5)}(bu) = \sum_{j=1}^{N_m} D(R_{:,j}) \cdot \left(bu - \left(\ln({R_{:,j}}^T) - \mu^T - {bm_j}^T - {M_{:,j}}^T \cdot U \cdot D\left(h(R_{:,j})\right)\right)^T\right)$$

The optimal solution is given by:

$$bu = \left(\sum_{j=1}^{N_m} D(R_{:,j})\right)^{-1} \cdot \sum_{j=1}^{N_m} D(R_{:,j}) \cdot \left(\ln({R_{:,j}}^T) - \mu^T - {bm_j}^T - {M_{:,j}}^T \cdot U \cdot D\left(h(R_{:,j})\right)\right)^T$$

The derivative w.r.t. μ is:

$$\frac{d}{d\mu} RMSE^{(5)}(\mu) = \sum_{i=1}^{N_u} D(R_{i,:}) \cdot \left(\mu - \left(\ln(R_{i,:}) - bm^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h(R_{i,:})\right)\right)^T\right)$$

The optimal solution is given by:

$$\mu = \left(\sum_{i=1}^{N_u} D(R_{i,:})\right)^{-1} \cdot \sum_{i=1}^{N_u} D(R_{i,:}) \cdot \left(\ln(R_{i,:}) - bm^T - {bu_i}^T - {U_{:,i}}^T \cdot M \cdot D\left(h(R_{i,:})\right)\right)^T$$

The penalization term and the friendship term can easily be added to the expressions for $U_{:,i}$ and $M_{:,j}$.

6. Matrix factorization with non-linear predictions

This problem can be formulated as:

$$RMSE^{(6)}(U,M) = \frac{1}{2}\sum_{i=1}^{N_u}\left(R_{i,:} - g\left(U_{:,i}{}^{T} \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)\right) \cdot \left(R_{i,:} - g\left(U_{:,i}{}^{T} \cdot M \cdot D\left(h\left(R_{i,:}\right)\right)\right)\right)^{T}$$

where $g(x)$ is a logistic function that gets vector as input and gives vector as output. The ratings should be normalized in the range $[0, 1]$. Penalization term and friendship term can also be added.