

3. Netflix Problem Report¹

Algorithm

The algorithm implemented for this assignment was **Alternating Least Squares** (ALS) with weighted- λ -regularization described in section 3.1 of the paper **Large-scale Parallel Collaborative Filtering for the Netflix Prize** by Zhou et al. The parallel approaches applied by Zhou et al. in sections 3.2 were not implemented in this project.

Zhou et al. formulate the matrix completion problem as a matrix-completion-by-factorization problem. They minimize the empirical, total loss which is simply the average of the square loss function defined as the squared error: $l^2(r, \mathbf{u}, \mathbf{m}) = (r - \langle \mathbf{u}, \mathbf{m} \rangle)^2$.

They regularize the empirical loss with a regularization term similar to Tikhonov regularization. The function is

$$f(U, M) = \sum_{(i,j) \in I} (r_{ij} - \mathbf{u}_i^T \mathbf{m}_j)^2 + \lambda \left(\sum_i n_{u_i} \|\mathbf{u}_i\|^2 + \sum_j n_{m_j} \|\mathbf{m}_j\|^2 \right)$$

They then derive $\mathbf{u}_i = A_i^{-1} V_i$ and $\mathbf{m}_j = A_j^{-1} V_j$ where $A_i = M_{I_i} M_{I_i}^T + \lambda n_{u_i} E$, $V_i = M_{I_i} R^T(i, I_i)$ and $A_j = U_{I_j} U_{I_j}^T + \lambda n_{m_j} E$, $V_j = U_{I_j} R^T(j, I_j)$ where E is the $r \times r$ identity matrix. See the paper for further details.

The pseudocode for the algorithm described in the paper and class notes can be summarized:

1. Initialize pattern matrix M by assigning the average rating for movies to the first row of M and small random numbers for the remaining entries. Initialize elements of U to small random numbers
2. Fix M , solve U by minimizing the objective function (sum of squared errors).
3. Fix U , solve M by minimizing the objective function.
4. Repeat Steps 2 and 3 until the algorithm has converged or a stopping criterion is satisfied.

Parameters and Results

Model parameters were λ and \mathbf{r} . Zhou et al.'s best reported scores (RMSE = 0.8985) were achieved with $\lambda = 0.065$ and $r = 1000$ after 50 iterations. The convergence criterion for this assignment was an iteration threshold $\mathbf{K} = 20$. Appropriate values for $\lambda = 0.1$ and $\mathbf{r} = 20$ were found by manually tuning on the test set.

I analyzed the results of running my algorithm on the provided data sets with Root Squared Mean Error (RSME) score. The results from this algorithm are inferior to those described in Zhou et al.'s paper with the Train-RMSE score converging to **0.69** and the Test-RMSE score going to approximately **0.958** after 20 iterations. I did not have enough time to rigorously optimize my parameters (λ, r) with statistical techniques (i.e. cross-validation), implement the global bias shift implemented in the original paper, or test the algorithm against a larger dataset - any of these factors may have contributed to the reduced RMSE scores compared to those reported by the paper.

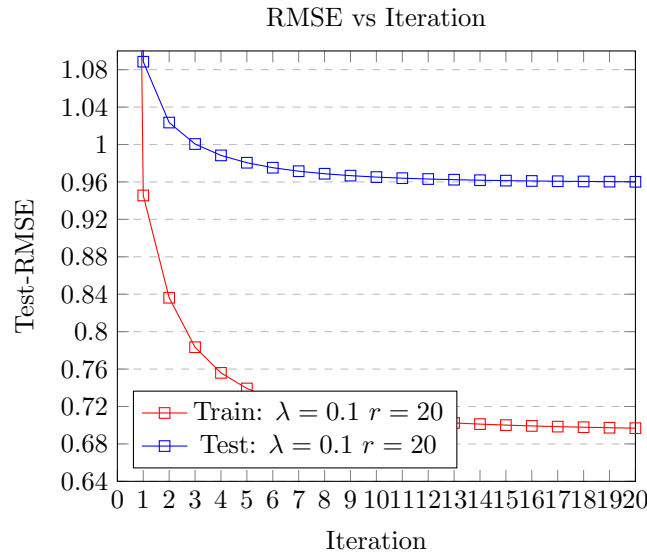


Figure 1: RMSE against Iteration

¹Large-Scale Parallel Collaborative Filtering for the Netflix Prize, Zhou et al.