Report on Visualization Project

Daiqi Linghu, Dunzhu Li, Yiran Ma

1 Introduction

2 Algorithm

2.1 Optimization

In the code (BSGD.m), we solve the following optimization problem:

$$\underset{U,V,a,b}{\operatorname{argmin}} \frac{\lambda}{2}\left(\|U\|^2+\|V\|^2+\|a\|^2+\|b\|^2\right)+\sum_{(i,j)\in S}(Q_{i,j})^2$$

where $Q_{i,j}=(u_i^T v_j+a_i+b_j+\mu-Y_{i,j})$. Then,

$$\frac{\partial J}{\partial U_l}=\lambda U_l+\sum_{(i,j)\in S}2Q_{i,j}1_{i=l}V_j$$

We call the first term in the gradient as "norm term", and the second term "error term".

U_l is updated iteratively through gradient method:

$$U_l^{n+1}=U_l^n-\eta\frac{\partial J}{\partial U_l}$$

In BSGD, for each iteration, we randomly loop through all the data and update U_l with "error term" only in the loop, then after the loop, correct U_l with "norm term". We can do V_k, a_l and b_k similarly.

2.2 Visualization

3 Implementation

3.1 Find U & V

To solve the optimization problem with Bias-SGD, we simply divide the data by 5 fold, use 4 folds as training data and 1 fold as test data to find the optimal damping parameter λ and number of iteration. The step size η is chosen as 1e-2 and is decreased by 10% for each iteration. From the results shown below (Fig. 1), we choose $\lambda = 25$. For $\lambda = 25$, the out-of-sample error is still decreasing at the 100th iteration (Fig. 1), however, the improvement is minimal. Therefore, we choose the number of iteration as 100.

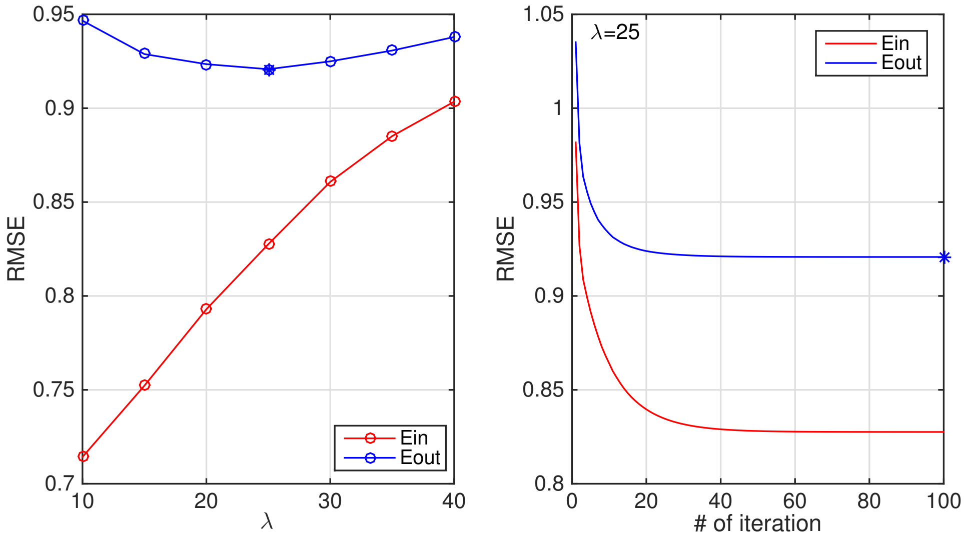

Figure 1: RMS error (RMSE) vs. Parameter. Left: For different λ, we plot the minimum RMSE in 100 iterations. We see that $\lambda = 25$ is the optimum choice. Right: For $\lambda = 25$, we plot the RMSEs in 100 iterations. We see that RMSE decays slowly after about 15 iterations. We simply choose the number of iteration as 100 for final run.

3.2 Projection