Daniel Alabi
Cody Wang **CS 324**

Regularized SVD

In class, we talked about the iterative SVD technique described in the book. We also talked briefly about how to avoid overfitting; one way to avoid overfitting is to use regularization [1].

For the netflix prize competition, Simon Funk implemented a regularized iterative SVD technique to predict movie ratings for users [2]. The goal of our project was to implement a regularized iterative SVD technique, similar to Simon Funk's. After Simon Funk wrote the article describing his SVD technique(s), a lot of other data scientists and researchers have tried to implement his approach. As a result, there is no homogenous regularized SVD technique. One thing that most of these techniques (including Simon Funk's) have in common is that they make use of a stochastic gradient descent technique to minimize the RMSE [3, 4]. Stochastic gradient descent is basically an iterative way to find a local minima of a function[5]. In our case, the function to be minimized is the RMSE.

Training the U, V matrices

The crux of the regularized SVD technique we implemented lies in how the U, V matrices are trained. Recall that, we are trying to "approximate" the utility matrix M with two matrices such that

$$M \approx U * V^T$$

Note that the singular values of the decomposition have been submerged into both U and V^T. Also, in the algorithm we implemented, instead of explicitly storing V^T, we store V to make the training step more seamless.

Training of the U and V^T matrices occurs exactly r times. r is the rank of the matrix $U * V^T$, or equivalently the number of columns in U, or the number of rows in V^T. For each k in $[0, r-1]$, we consider

Training minimizes train RMSE – that is, ... We hope it also minimizes test RMSE

Varying the Parameters

Control Experiments

Method of Research and Future Work

1. Cache.
2. Average ratings.

References

1. http://en.wikipedia.org/wiki/Regularization_(mathematics)
2. http://sifter.org/ simon/Journal/20061211.html
3. http://www.timelydevelopment.com/demos/NetflixPrize.aspx
4. http://alias-i.com/lingpipe/docs/api/com/aliasi/matrix/SvdMatrix.html
5. http://en.wikipedia.org/wiki/Stochastic_gradient_descent