Understanding Recommendation Systems

Indiana University

February 22, 2010

Recommendation Systems

Netflix and **goodreads.com** allow users to rate movies and books.

We have ratings for *user / item* pairs (e.g., *Eric* rated *The Departed* a 5).

Clearly, we do not know ratings for items the user has not rated.

We could recommend an item if we believe a user would rate it highly. Therefore, the ability to predict a rating that a user would give an item can be used for recommendation systems.

Recommendation Systems

We collect the known ratings into a matrix $\widehat{M}$.

Movies

	3			2			4			5	1				
					2		3	5		4		3		1	
				3				5			1				
		1		2		4			3		4		5		
			4		5		2			1		4			
Users					5	1		4	3		2		3		2
	1		3		2			4	5			4			2
		2	2		3		5		4	3		2	1		
			2			2		3		4		3	4		
	2		3	4		4	2				3	4			

We assume there is a **true matrix** M that is completely filled with the ratings every user would give each item.

Recommendation Systems

Netflix hosted a \$1 million prize to a team that could make a substantial improvement in accuracy over their method to fill in the "missing entries". They withheld known entries to measure how well participants' did.

A number of researchers have developed new methodology to make such predictions. One popular method is the so-called low rank matrix completion problem. Here, we assume M is low rank and try to approximate it based only on the known entries in $\widehat{M}$ such that

$$M \approx UV^T$$

where the matrices U and V are low rank and UV^T approximates the known entries $\widehat{M}$.

Recommendation Systems

This project will be two-fold.

1. We will implement a low rank matrix completion system for predicting user ratings on books. It will have a web front end and allow users to enter a handful of ratings and return book recommendations. The training data will consist of data from a portion of the *Book-Crossing* user/book rating dataset (http://www.informatik.uni-freiburg.de/ cziegler/BX).

2. Additionally, we are interested in studying the effect of rounding ratings. We assume that users truly rate movies with real numbers between 1 and 5 (e.g., 3.77), but are forced to round to the nearest integer (viz., $1, 2, 3, 4,$ or 5). We see rounding as inducing error and are interested in studying its effect on the low rank matrix approximation.