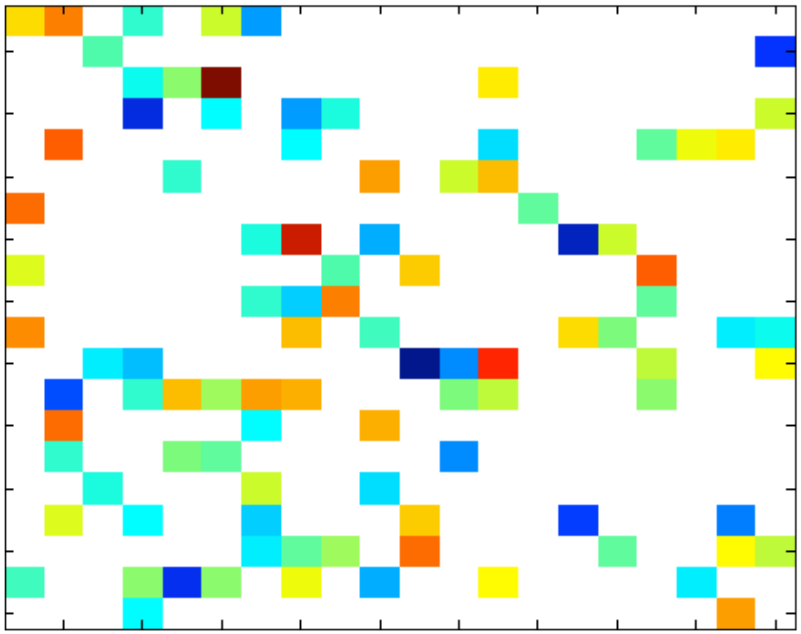
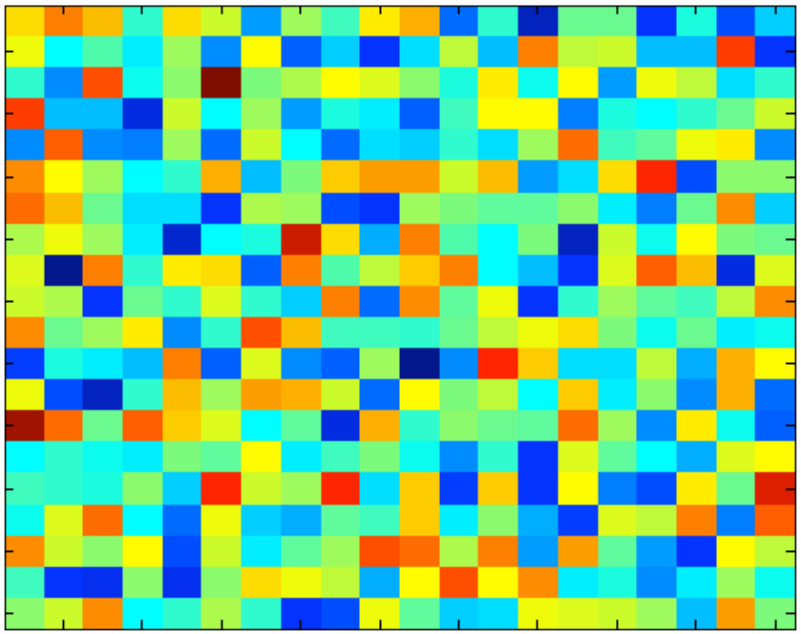
Unsupervised Learning - Matrix Completion

Creating our own truth





amazon.com

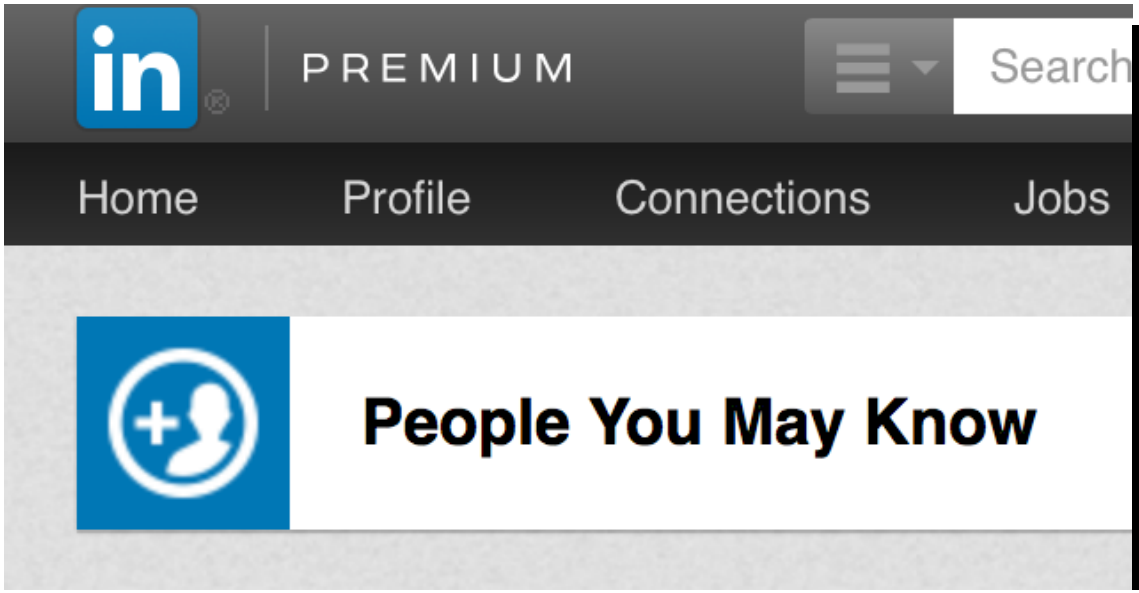
Recommended for You

Amazon.com has new recommendations for you based on items you purchased or told us you own.

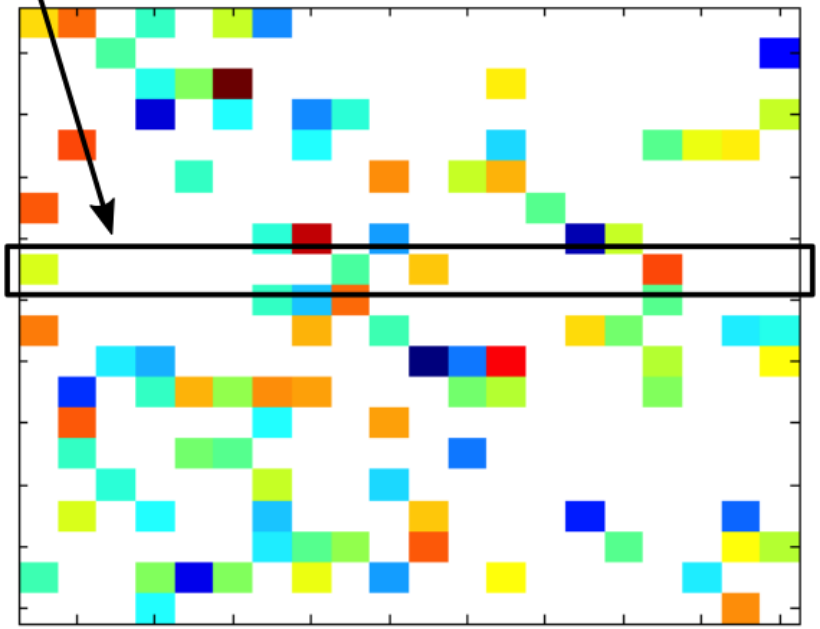
Google Apps Deciphered: Compute in the Cloud to Streamline Your Desktop

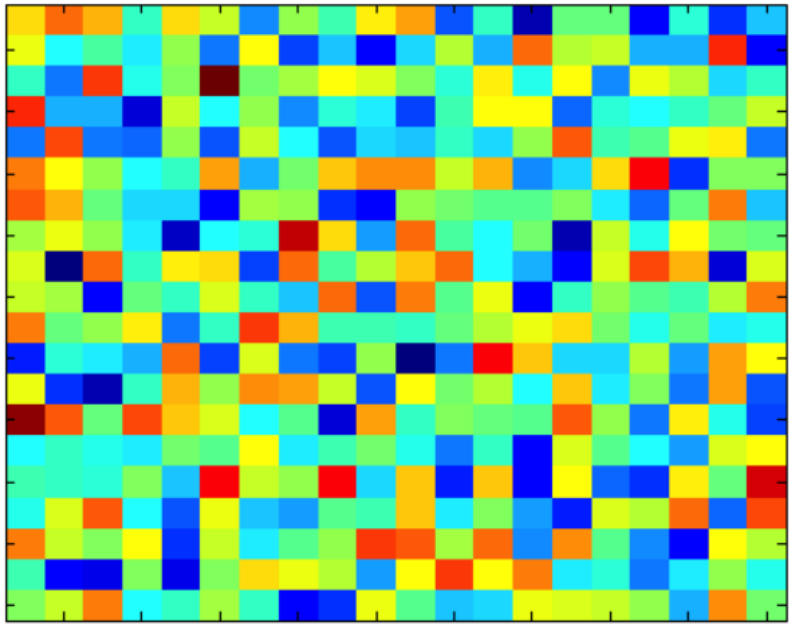
Google Apps Administrator Guide: A Private-Label Web Workspace

Googlepedia: The Ultimate Google Resource (3rd Edition)



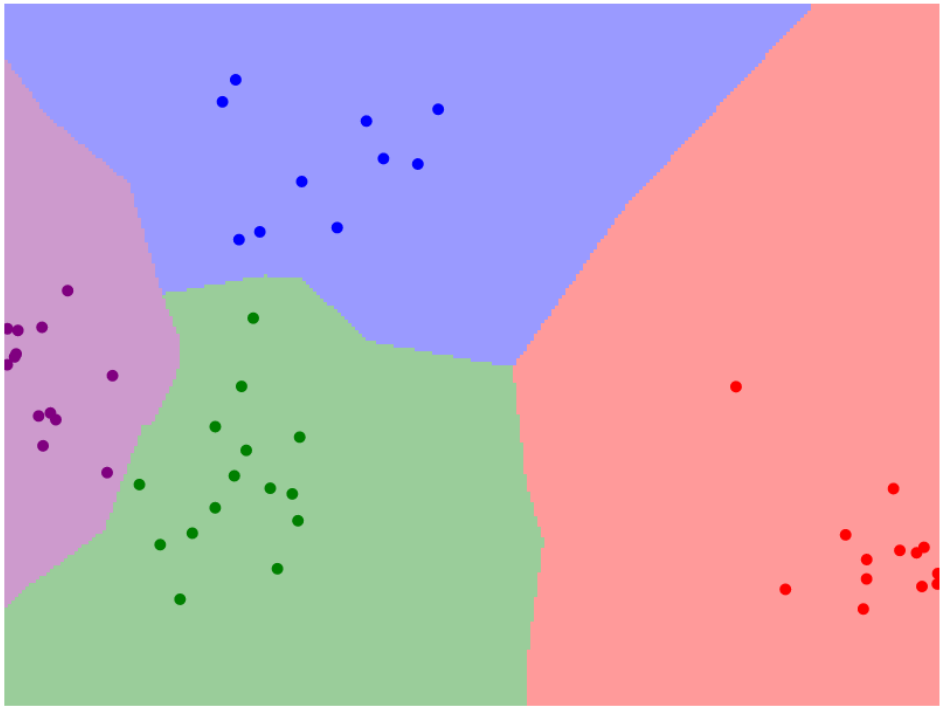
You (a single row)





How Do I Want to Fill in the Missing Entries?

k-Nearest Neighbors



Assumption: My "suggested friends" should be an average of my *k* closest friends' lists.

k-Nearest Neighbors

$$D_{euclidian} = \left(\sum_{i=i}^{n} |x_i - y_i|^2\right)^{1/2}$$

$$D_{manhattan} = \sum_{i=i}^{n} |x_i - y_i|$$

$$D_{euclidian} = \sqrt{1^2 + 1^2 + 1^2 + 1^2 + 10^2} = \sqrt{104} \approx 10$$

$$D_{manhattan} = 1 + 1 + 1 + 1 + 10 = 14$$

Low Rank Matrix Completion

$$\begin{aligned} \min_{M} \quad & \text{rank}(M) \\ \text{subject to} \quad & M_{ij} = X_{ij} \end{aligned}$$

The *rank* of a matrix is defined as the number of linearly independent column vectors in a matrix.

Assumption: Our data matrix X can be represented by a small number of characteristics.

Low Rank Matrix Completion

		-1		
			1	
1	1	-1	1	-1
1				-1
		-1		

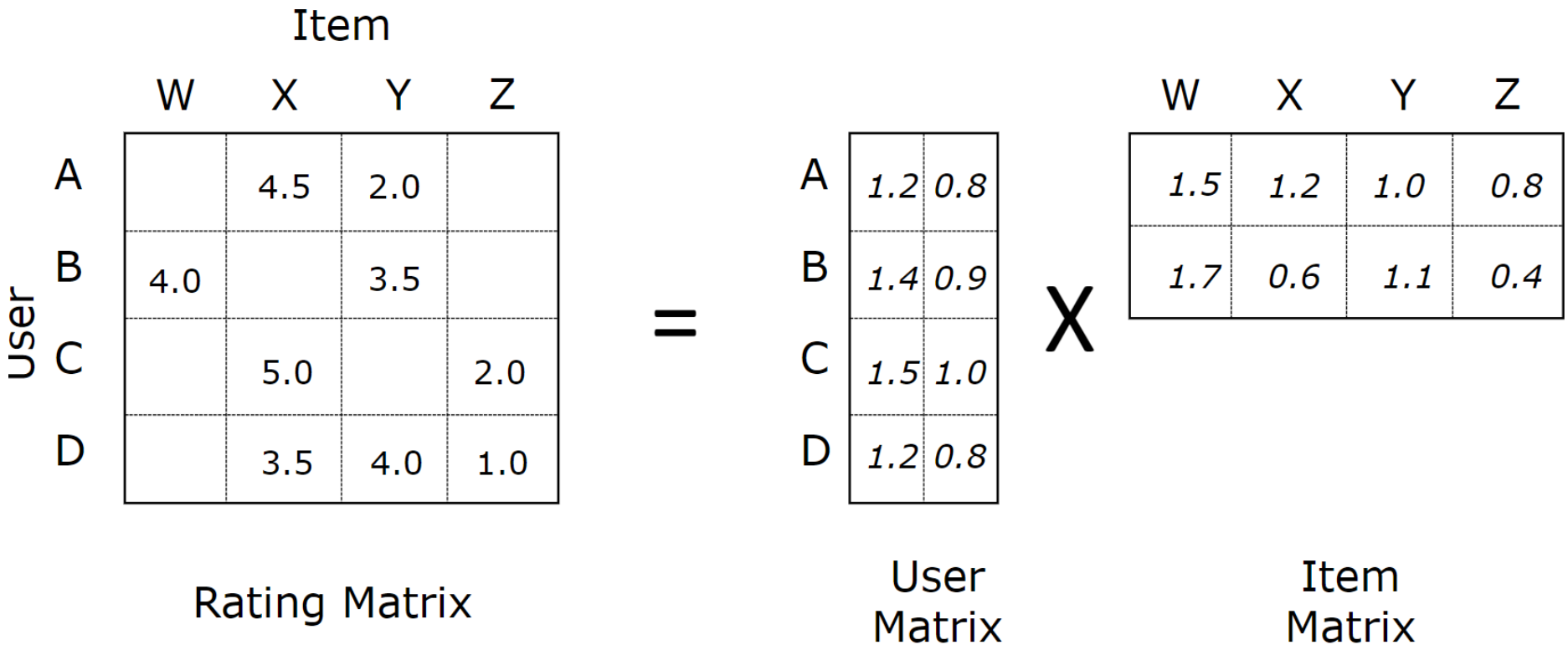
1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1
1	1	-1	1	-1

Matrix Factorization

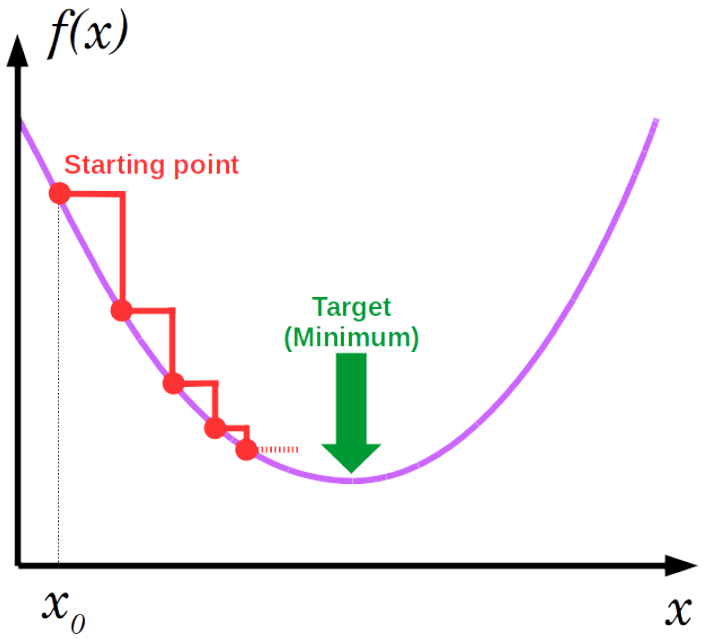
$$\min_{P,Q} \quad ||X - PQ||_F^2$$

$$\text{where} \quad X \in \mathbb{R}^{n \times d},\ P \in \mathbb{R}^{n \times r},\ Q \in \mathbb{R}^{r \times d},$$

$$r << d$$



Matrix Factorization - Gradient Descent



Drawback: Since $P \in \mathbb{R}^{n \times r}$ and $Q \in \mathbb{R}^{r \times d}$ there are $(n \times r) + (r \times d)$ variables to optimize.

Matrix Factorization - Alternating Minimization

The Idea:

1. Fix Q and update P.
2. Then, fix P and update Q.
3. Rinse and repeat.

$$\min_{P,Q} \quad ||X - PQ||_F^2$$

Derived Updates:

$$P^{t+1} = XQ^T(QQ^T)^{-1}$$
$$Q^{t+1} = (P^TP)^{-1}P^TX$$

Questions

These slides are designed for educational purposes, specifically the CSCI-470 Introduction to Machine Learning course at the Colorado School of Mines as part of the Department of Computer Science.