# Project Plan

## Zack Tillotson

## Oct 2010

Nonnegative matrix factorization (NMF) is a technique which has gained prominence in the last decade as a powerful machine learning tool.

In NMF a sparse input matrix S is factored into two matrices A and B such that S = AB. These two "feature" matrices attempt to encode the patterns of the original matrix, such that when A and B are expanded into a new matrix S' which approximates S, the blank elements from S now have values in S' to be used in classification. The results gained from this type of algorithm are generally of very good quality and it has also been applied to very large data sets.

NMF is simply a grand strategy for solving the problems, how to actually decompose the source matrix in the feature matrices is still of debate. There have been many algorithms offered, including Multiplicative Update (MU), Singular Value Decomposition (SVD), and Principal Component Analysis (PCA).

I plan to focus my research on the different algorithms for creating the decomposition and compare them to each other based on computational complexity, quality of the solution, and practical performance. There are several large data sets I can work with to do these tests, including the Netflix Competition's test set, which consists of over 100 million ratings that nearly 500,000 users gave to 15,000 movies. The very large size of this data set will allow me to strain the implementations of the algorithms and provide quality results.

I feel that presenting an analysis of the different algorithmic approaches to NMF will facilitate the understanding of NMF itself and will contribute to the understanding of the practical aspects of applying of machine learning algorithms to real world problems.