rIBP Optimization Formulation

Due: 2016/03/03

Collaborators:
Taylor Killian,
Leonhard Spiegelberg

- Given data $\mathcal{X}$ with dimensions (N $\times$ D)
- Find a factorization $\mathcal{W}\mathcal{A}$, where $\mathcal{W}$ is N by K and $\mathcal{A}$ is K by D that satisfies the following optimization problem:

$$\min \|\mathcal{X} - \mathcal{W}\mathcal{A}\|_F$$

$$\text{s.t } \mathcal{W}_{nk} \in \{0, 1\}$$
$$\sum_k \mathcal{W}_{nk} = c_n$$

The objective in the above formulation encourages the solution to satisfy $\mathcal{X} \approx \mathcal{W}\mathcal{A}$ where (at this initial stage) the entries of $\mathcal{A}$ are unconstrained. There are two simple relaxations that make the formation of $\mathcal{W}$ a little easier. From discussions with Finale, it seemed that part is the hardest. The first relaxation to relax the first constraint is to label the entries of $\mathcal{W}$ with the following distribution:

$$\mathcal{W}_{nk} = \begin{cases} 0 & \text{with prob. } p \\ \text{some } w \in \mathbb{R} & \text{with prob. } 1 - p \end{cases}$$

The second relaxation, simplifying the second constraint, is to create a penalty/loss term, based on the distance from c_n the row $\mathcal{W}_n$ is, in the objective of the form:

$$(c_n - \sum_k \mathcal{W}_{nk})^2$$