Feng Liang
Visible to: Everyone

Project 2: What We Have Tried (III) #457

Project 2

As illustrated in the subsequent figures, the sales trends for a department appear consistent across different stores. To diminish noise, we employ Singular Value Decomposition (SVD) as detailed below:

- Organize data from a specific department into a matrix $X_{m \times n}$ where m represents the number of stores that have this particular department and n is the number of weeks.
- You may encounter some missing values in the matrix X. You'll need to select a method for handling these missing values. In my approach, for simplicity, I chose to replace all missing values with zero.
- Implement SVD (or PCA, with observations as stores and weeks as features):
  $X - \text{store.mean} = U_{m \times r} D_{r \times r} V^t_{r \times n}$
  where $r \leq \min(n, m)$ denotes the rank of X, and D is a diagonal matrix.
- Choose the top d components, and obtain a reduced rank (or smoothed) version of the original dataset:
  $\tilde{X} = U_{m \times r} \tilde{D}_{r \times r} V^t_{r \times n} + \text{store.mean}$
  where $\tilde{D}_{r \times r}$ is a diagonal matrix keeping its first d diagonal values from D and setting the subsequent diagonal values to zero.
- Of course, if the rank of X is less than or equal to d, $\tilde{X}$ would be the same as X. For example, if the rows of X (i.e., the number of stores) is less than or equal to d, you can skip the SVD part.
- Alternatively, we can compute $\tilde{X}$ by preserving only the elements in U, D, V related to the top d components:
  $\tilde{X} = U[, 1:d] \times D[1:d, 1:d] \times (V[, 1:d])^t + \text{store.mean}$