Recommendations Using Coupled Matrix Factorization

Abhinav Jauhri

November 12, 2014

Backup Slides

NNMF - Forbenius norm minimization

Theorem

The Euclidean distance $\|V - WH\|_F$ is nonincreasing under the update rules:

$$H_{au} \leftarrow H_{au} \sum_i W_{ia} \frac{V_{iu}}{(WH)_{iu}} \tag{1}$$

$$W_{ia} \leftarrow W_{ia} \sum_u \frac{V_{iu}}{(WH)_{iu}} H_{au} \tag{2}$$

Idea:

- if ratio > 1, then we need to increase denominator
- if ratio < 1, then decrease the denominator
- if ratio $= 1$, do nothing

NNMF - KL Divergence minimization

Theorem

Let $D(A\|B) = \sum_{ij}(A_{ij} \log \frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij})$.

The divergence $D(V\|WH)$ *is nonincreasing under similar update rules.*

Ridge Regression is biased

- Let $R = A^T A$
-

$$\begin{aligned}
\beta_\lambda^{ridge} &= (A^T A + \lambda I_k)^{-1} A^T y \\
&= (R + \lambda I_k)^{-1} R (R^{-1} A^T y) \\
&= [R(I_k + \lambda R^{-1})]^{-1} R[(A^T A)^{-1} A^T y] \\
&= (I_k + \lambda R^{-1})^{-1} R^{-1} R \beta^{ls} \\
&= (I_k + \lambda R^{-1}) \beta^{ls}
\end{aligned} \tag{3}$$

- Therefore,

$$\begin{aligned}
\mathbb{E}(\beta_\lambda^{ridge}) &= \mathbb{E}((I_k + \lambda R^{-1}) \beta^{ls}) \\
&\neq \beta^{ls} \text{ if} \lambda \neq 0
\end{aligned} \tag{4}$$

Computational Complexities

- For k features and n training examples, CMF - $O(k^2 n)$
  - $X^T X$ - $O(k^2 n)$
  - $X^T y$ - $O(kn)$
  - $(X^T X)^{-1} X^T y$ + factorization - $O(k^3)$

  Assume $n > k$, otherwise $X^T X$ would be singular (and hence non-invertible). Therefore $O(k^2 n)$
- Singular Value Decomposition - $O(min(mn^2, m^2 n))$, for a $m \times n$ matrix
- NNMF - Computational complexity for each iteration is $O(knm)$

Forbenius is equivalent to spectral - SVD

We have that $\|A - A_k\|_F = \|U_{k+1:r}\Sigma_{k+1:r}V_{k+1:r}^T\|_F = \sqrt{\sum_{i=k+1}^r \sigma_i^2}$.
Now let $A - C$ have svd $\widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$. Then $\|A - C\|_F^2 = \sum_{i=1}^n \widetilde{\sigma}_i{}^2$.
Thus, we can write:

$$\begin{aligned} \|A - C\|_F^2 &= \widetilde{\sigma_1^2} + \cdots + \widetilde{\sigma_n^2} \\ &= \|A - C\|_2^2 + \|(A - C) - \widetilde{\sigma_1}\widetilde{u_1}\widetilde{v_1^T}\|_2^2 + \cdots + \|(A - C) - \sum_{i=1}^{n-1}\widetilde{\sigma_i}\widetilde{u_i}\widetilde{v_i^T}\|_2^2 \end{aligned} \tag{5}$$

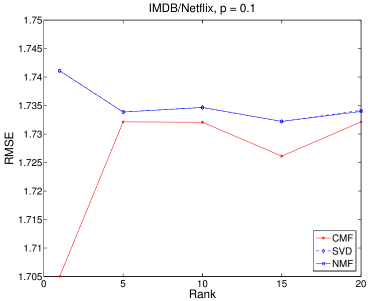
But $C + \sum_{i=1}^{l}\widetilde{\sigma_i}\widetilde{u_i}\widetilde{v_i^T}$ has rank at most $k + l$

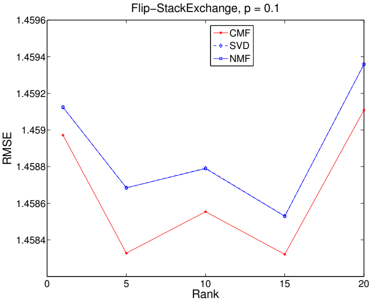
$$\left\|A - (C + \sum_{i=1}^{l}\sigma_i u_i v_i^T)\right\|_2 \geq \|A - A_{k+l}\|_2 = \sigma_{k+l+1} \tag{6}$$

Applying (6) to (5):

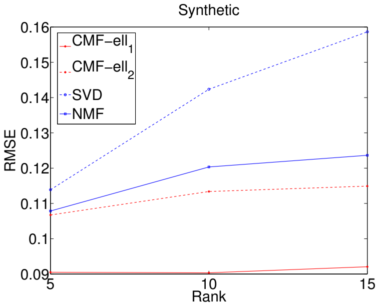
$$\|A - C\|_F^2 \geq \sum_{i=k}^{k+n+1}\sigma_{i+1}^2 = \sum_{j=k+1}^{r}\sigma_j^2 = \|A - A_k\|_F^2 \tag{7}$$

Flip Results





Completion RMSE for 10% missing values



Future Work - Canonical Correlation Analysis

- Attempts to answer - *which direction accounts for much of the covariance between two data sets*[1]
- Given $X \in \mathbb{R}^{n \times p}$, $Y \in \mathbb{R}^{n \times q}$, we get two directions $\alpha_1 \in \mathbb{R}^p, \beta_1 \in \mathbb{R}^q$ that maximizes sample covariance:

$$\alpha_1, \beta_1 = \underset{\|X\alpha\|_2=1, \|Y\beta\|_2=1}{\operatorname{argmax}} cov(X\alpha, Y\beta) \tag{8}$$

[1]Source: 36-462, Data Mining by Ryan Tibshirani