COL866: Foundations of Data Science

Ragesh Jaiswal, IITD

Power Method for Singular Value Decomposition (SVD)

- Let $B = A^T A$
- Question: Can you point out some interesting properties of B?

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$
- Question: Can we obtain a similar expression for B^2 and in general B^k?

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$
- Question: Can we obtain a similar expression for B^2 and in general B^k?
- $B^k = \sum_{i=1}^{r} \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T$
- So, if $\sigma_1 > \sigma_2$, then normalizing the first column of B^k should give a good estimate for $\mathbf{v}_1$.

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$
- Question: Can we obtain a similar expression for B^2 and in general B^k?
- $B^k = \sum_{i=1}^{r} \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T$
- So, if $\sigma_1 > \sigma_2$, then normalizing the first column of B^k should give a good estimate for $\mathbf{v}_1$.
- A faster method:
  - Computing B^k may be costly.
  - Select a random vector $\mathbf{x} = \sum_{i=1}^{d} c_i \mathbf{v}_i$.
  - Claim: $B^k \mathbf{x} \approx \sigma_1^{2k} c_1 \mathbf{v}_1$
  - So, normalizing $B^k \mathbf{x}$ approximates $\mathbf{v}_1$.

- A faster method:
  - Computing B^k may be costly.
  - Select a random vector $\mathbf{x} = \sum_{i=1}^{d} c_i \mathbf{v}_i$.
  - Claim: $B^k \mathbf{x} \approx \sigma_1^{2k} c_1 \mathbf{v}_1$
  - So, normalizing $B^k \mathbf{x}$ approximates $\mathbf{v}_1$.
- Given $\min_{i<j} \log\left(\frac{\sigma_i}{\sigma_j}\right) \geq \lambda$, the following algorithm estimates (within ε error with probability $\geq (1-\delta)$) the first singular value and singular vectors.

Algorithm

1. Generate $\mathbf{x}_0$ from a spherical gaussian with mean 0 and variance 1.
2. $s \leftarrow \log\left(\frac{8d \log(2d/\delta)}{\varepsilon\delta}\right)/2\lambda$
3. For $i = 1$ to s
4. $\quad \mathbf{x}_i \leftarrow (A^T A)\mathbf{x}_{i-1}$
5. $\mathbf{v}_1 \leftarrow \mathbf{x}_i / ||\mathbf{x}_i||$
6. $\sigma_1 \leftarrow ||A\mathbf{v}_1||$
7. $\mathbf{u}_1 \leftarrow A\mathbf{v}_1/\sigma_1$
8. return$(\sigma_1, \mathbf{u}_1, \mathbf{v}_1)$

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$
- Question: Can we obtain a similar expression for B^2 and in general B^k?
- $B^k = \sum_{i=1}^{r} \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T$
- So, if $\sigma_1 > \sigma_2$, then normalizing the first column of B^k should give a good estimate for $\mathbf{v}_1$.
- A faster method:
  - Computing B^k may be costly.
  - Select a random vector $\mathbf{x} = \sum_{i=1}^{d} c_i \mathbf{v}_i$.
  - Claim: $B^k \mathbf{x} \approx \sigma_1^{2k} c_1 \mathbf{v}_1$
  - So, normalizing $B^k \mathbf{x}$ approximates $\mathbf{v}_1$.
- The above approximations are with respect to the fact that σ_1 is significantly larger than σ_2. What if this is not true?

- Let $B = A^T A$
- $B = \sum_{i=1}^{r} \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^T$
- Question: Can we obtain a similar expression for B^2 and in general B^k?
- $B^k = \sum_{i=1}^{r} \sigma_i^{2k} \mathbf{v}_i \mathbf{v}_i^T$
- So, if $\sigma_1 > \sigma_2$, then normalizing the first column of B^k should give a good estimate for $\mathbf{v}_1$.
- A faster method:
  - Computing B^k may be costly.
  - Select a random vector $\mathbf{x} = \sum_{i=1}^{d} c_i \mathbf{v}_i$.
  - Claim: $B^k \mathbf{x} \approx \sigma_1^{2k} c_1 \mathbf{v}_1$
  - So, normalizing $B^k \mathbf{x}$ approximates $\mathbf{v}_1$.
- The above approximations are with respect to the fact that σ_1 is significantly larger than σ_2. What if this is not true?

Theorem

Let A be an $n \times d$ matrix and $\mathbf{x}$ a unit length vector in $\mathbb{R}^d$ with $\mathbf{x}_t \mathbf{v}_1 \geq \delta$, where $\delta > 0$. Let V be the space spanned by the right singular vectors of A corresponding to singular values greater than $(1 - \epsilon)\sigma_1$. Let $\mathbf{w}$ be the unit vector after $k = \frac{\ln 1/\epsilon\delta}{2\epsilon}$ iterations of the power method, namely $\mathbf{w} = \frac{(A^T A)^k \mathbf{x}}{||(A^T A)^k \mathbf{x}||}$. Then $\mathbf{w}$ has a component of at most ϵ perpendicular to V.

End