TTIC 31150/CMSC 31150
Mathematical Toolkit (Spring 2023)

Avrim Blum and Ali Vakilian

Lecture 8: SVD applications

Recap

- SVD: Let $\sigma_1^2 \geq \cdots \geq \sigma_r^2 > 0$ be nonzero eigenvalues of $\varphi^*\varphi$ with corresponding orthonormal eigenvectors $v_1, \ldots, v_r$. Let $w_i = \varphi(v_i)/\sigma_i$. Then:
  - $w_1, \ldots, w_r$ are orthonormal, $\varphi(v_i) = \sigma_i w_i$ and $\varphi_i^*(w_i) = \sigma_i v_i$.
  - $\varphi = \sum_{i=1}^{r} \sigma_i \, |w_i\rangle\langle v_i|$, where $|w_i\rangle\langle v_i|$ is outer product.

- Matrix view: $A = \sum_{i=1}^{r} \sigma_i w_i v_i^* = W\Sigma V^*$, where W has $w_1, \ldots, w_r$ as columns, V^* has $v_1^*, \ldots, v_r^*$ as rows, and Σ is an $r \times r$ diagonal matrix with $\Sigma_{ii} = \sigma_i$.

- Let $A_k = \sum_{i=1}^{k} \sigma_i w_i v_i^*$. Then:

$$\|(A-B)\|_2 = \max_{v \neq 0} \frac{\|(A-B)v\|_2}{\|v\|_2}.$$

Proposition 2.1 $\|A - A_k\|_2 = \sigma_{k+1}$.

Proposition 2.4 *Let* $B \in \mathbb{C}^{m \times n}$ *have* $\mathrm{rank}(B) \leq k$ *and let* $k < r$. *Then* $\|A - B\|_2 \geq \sigma_{k+1}$.

Frobenius norm approximation

Now: show that A_k is the rank-k matrix B minimizing Frobenius norm $\sqrt{\sum_{ij}(A-B)_{ij}{}^2}$.

Equivalently, if we think of each row of A as a data point, we are finding the rank-k subspace that minimizes the mean squared distance of the points to that subspace (A_k represents projecting each point in A to this subspace).

Will use this view in our discussion.

To match the notes, $m \to n, n \to d$.

Least squares approximation

Let $a_1, \ldots, a_n \in \mathbb{R}^d$. Want to find subspace S that minimizes $\sum_{i=1}^{n} dist(a_i, S)^2$.

Claim 1.1 *Let $u_1, \ldots, u_k$ be an orthonormal basis for S. Then*

$$(\mathrm{dist}(a_i, S))^2 = \|a_i\|_2^2 - \sum_{j=1}^{k} \langle a_i, u_j \rangle^2 .$$

Remark:

- The $\boldsymbol{dist(a_i, S)}$ is independent of choice of orthonormal basis.
    - *Different ways of computing it but not different quantities*

Least squares approximation

Let $a_1, \dots, a_n \in \mathbb{R}^d$. Want to find subspace S that minimizes $\sum_{i=1}^{n} dist(a_i, S)^2$.

Claim 1.1 *Let $u_1, \dots, u_k$ be an orthonormal basis for S. Then*

$$(\mathrm{dist}(a_i, S))^2 = \|a_i\|_2^2 - \sum_{j=1}^{k} \langle a_i, u_j \rangle^2 .$$

Proof:

- Can write $a_i = a_i^{in} + a_i^{perp}$ where a_i^{in} is the projection of a_i to S and a_i^{perp} is orthogonal to S.
- Get $\left\|a_i^{perp}\right\|^2 = \|a_i\|^2 - \left\|a_i^{in}\right\|^2$.
- Formally, extending $u_1, \dots, u_k$ to orthonormal basis for $\mathbb{R}^d$ and writing a_i in this basis.

Computing $dist(a_i, S)$

$u_1, \cdots, u_k, \qquad u_{k+1}, \cdots, u_d$

- Any $u \in S$ can be written as $u \coloneqq \sum_{j=1}^{k} b_j\, u_j$
- $a_i \coloneqq \sum_{j=1}^{d} c_j\, u_j$
- $a_i - u = \sum_{j=1}^{k} (c_j - b_j)\, u_j + \sum_{j=k+1}^{d} c_j\, u_j$

- $\|a_i - u\|^2 = \sum_{j=1}^{k} |c_j - b_j|^2 + \sum_{j=k+1}^{d} |c_j|^2$

$$\geq \sum_{j=k+1}^{d} |c_j|^2$$

$$= \sum_{j=1}^{d} |c_j|^2 - \sum_{j=1}^{k} |c_j|^2 = \|a_i\|^2 - \sum_{j=1}^{k} |\langle a_i, u_j \rangle|^2$$

Least squares approximation

Let $a_1, \ldots, a_n \in \mathbb{R}^d$. Want to find subspace S that minimizes $\sum_{i=1}^{n} dist(a_i, S)^2$.

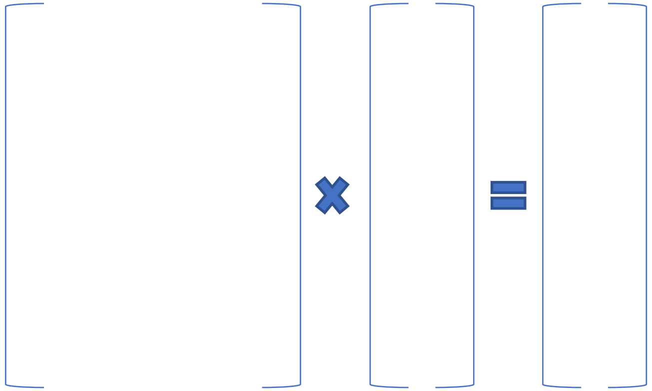
Claim 1.1 *Let $u_1, \ldots, u_k$ be an orthonormal basis for S. Then*

$$(\text{dist}(a_i, S))^2 = \|a_i\|_2^2 - \sum_{j=1}^{k} \langle a_i, u_j \rangle^2 .$$

Since the 1st term on the RHS is fixed, our goal can be viewed as: find k orthonormal vectors $u_1, \ldots, u_k$ to maximize $\sum_{i=1}^{n} \sum_{j=1}^{k} \langle a_i, u_j \rangle^2$.

Equivalently (with A as the matrix with a_i^T as row i),

we want to maximize $\sum_{j=1}^{k} \sum_{i=1}^{n} \langle a_i, u_j \rangle^2 = \sum_{j=1}^{k} \|Au_j\|^2$.



Least squares approximation

Proposition 1.2 *Let $v_1, \ldots, v_r$ be the right singular vectors of A corresponding to singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Then, for all $k \leq r$ and all orthonormal sets of vectors $u_1, \ldots, u_k$*

$$\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2 \leq \|Av_1\|_2^2 + \cdots + \|Av_k\|_2^2$$

Proof: by induction on k.

Base case ($k = 1$):

- $\|Au_1\|^2 = \langle Au_1, Au_1 \rangle = \left\langle u_1, A^T A u_1 \right\rangle \leq \max_{v \in \mathbb{R}^d} \mathcal{R}_{A^T A}(v) = \sigma_1^2 = \|Av_1\|^2.$

Least squares approximation

Proposition 1.2 *Let $v_1, \dots, v_r$ be the right singular vectors of A corresponding to singular values $\sigma_1 \geq \cdots \geq \sigma_r > 0$. Then, for all $k \leq r$ and all orthonormal sets of vectors $u_1, \dots, u_k$*

$$\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2 \leq \|Av_1\|_2^2 + \cdots + \|Av_k\|_2^2$$

Proof: by induction on k.

General k:

- Let's define $V_{k-1}^{\perp} = \{v \in \mathbb{R}^d : \langle v, v_i \rangle = 0 \;\; \forall i \in \{1, \dots, k-1\}\}$, and assume for now that $u_k \in V_{k-1}^{\perp}$.
- So, $\|Au_k\|^2 \leq \max_{v \in V_{k-1}^{\perp}, \|v\|=1} \|Av\|^2 = \sigma_k^2 = \|Av_k\|^2$.
- And $\|Au_1\|^2 + \cdots + \|Au_{k-1}\|^2 \leq \|Av_1\|^2 + \cdots + \|Av_{k-1}\|^2$ by induction. So, done.

So, just need to argue why we can assume wlog that $u_k \in V_{k-1}^{\perp}$.

Least squares approximation

Claim 1.3 *Given an orthonormal set $u_1, \ldots, u_k$, there exist orthonormal vectors $u'_1, \ldots, u'_k$ such that*

- $u'_k \in V_{k-1}^{\perp}$.
- $\text{Span}\,(u_1, \ldots, u_k) \;=\; \text{Span}\,(u'_1, \ldots, u'_k)$.
- $\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2 \;=\; \|Au'_1\|_2^2 + \cdots + \|Au'_k\|_2^2$.

Proof (similar to a proof we used last class):

- Since $\dim(V_{k-1}^{\perp}) = d - k + 1$ and $\dim(\text{Span}(u_1, \ldots, u_k)) = k$, there must exist some u'_k in the intersection with $\|u'_k\| = 1$.
- Complete to an orthonormal basis $u'_1, \ldots, u'_k$ of $\text{Span}(u_1, \ldots, u_k)$.
- Satisfies 3rd property because LHS and RHS both equal the sum of squared lengths of the projections of the rows of A into this k-dimensional subspace.

Gershgorin Disc Theorem

Sum of absolute values of off-diagonal entries in row i.

Theorem 2.1 (Gershgorin Disc Theorem) *Let* $M \in \mathbb{C}^{n\times n}$. *Let* $R_i = \sum_{j\neq i} |M_{ij}|$. *Define the set*

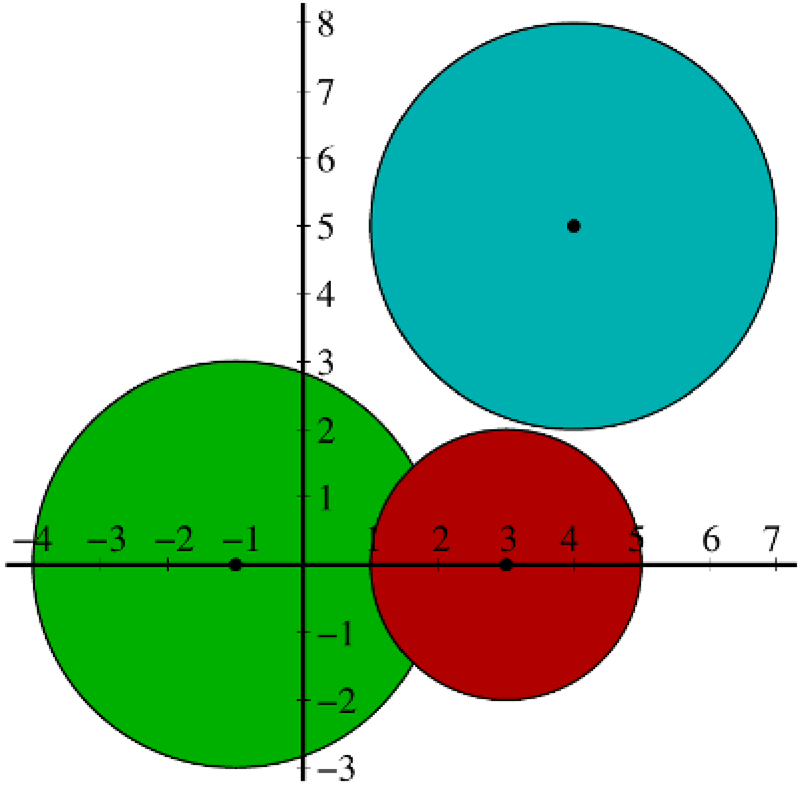
$$\mathrm{Disc}(M_{ii}, R_i) := \{z \in \mathbb{C} : |z - M_{ii}| \leq R_i\} .$$

If λ *is an eigenvalue of* M, *then*

$$\lambda \in \bigcup_{i=1}^{n} \mathrm{Disc}(M_{ii}, R_i) .$$

If matrix is close to being diagonal,

then eigenvalues are close to the diagonal entries.



source: golem.ph.utexas.edu

Gershgorin Disc Theorem

Sum of absolute values of off-diagonal entries in row i.

Theorem 2.1 (Gershgorin Disc Theorem) *Let* $M \in \mathbb{C}^{n\times n}$. *Let* $R_i = \sum_{j\neq i} |M_{ij}|$. *Define the set*

$$\mathrm{Disc}(M_{ii}, R_i) := \{z \in \mathbb{C} : |z - M_{ii}| \leq R_i\} .$$

If λ *is an eigenvalue of* M, *then*

$$\lambda \in \bigcup_{i=1}^{n} \mathrm{Disc}(M_{ii}, R_i) .$$

Input:

eigenvalues	$\begin{pmatrix} 5 & 0.1 & 0.1 \\ -0.1 & 6 & 0.1 \\ 0.1 & 0.1 & 7 \end{pmatrix}$

Results:

$\lambda_1 \approx 7.01475$

$\lambda_2 \approx 5.98019$

$\lambda_3 \approx 5.00506$

If matrix was perfectly diagonal, then eigenvalues would be exactly the diagonal entries.

Proof strategy: for eigenvector x, pick coordinate x_{i_0} of largest absolute value. Show eigenvalue close to $M_{i_0 i_0}$.

Gershgorin Disc Theorem

Sum of absolute values of off-diagonal entries in row i.

Theorem 2.1 (Gershgorin Disc Theorem) *Let* $M \in \mathbb{C}^{n \times n}$*. Let* $R_i = \sum_{j \neq i} |M_{ij}|$*. Define the set*

$$\mathrm{Disc}(M_{ii}, R_i) := \{z \in \mathbb{C} : |z - M_{ii}| \leq R_i\} .$$

If λ *is an eigenvalue of* M*, then*

$$\lambda \in \bigcup_{i=1}^{n} \mathrm{Disc}(M_{ii}, R_i) .$$

Proof:

- Let x be an eigenvector with eigenvalue λ. Let x_{i_0} be coordinate of largest absolute value.
- $\sum_j M_{i_0 j} x_j = \lambda x_{i_0}$. So, $\sum_{j \neq i_0} M_{i_0 j} x_j = \lambda x_{i_0} - M_{i_0 i_0} x_{i_0}$.
- So, $|\lambda - M_{i_0 i_0}| \leq \sum_{j \neq i_0} \frac{|M_{i_0 j}||x_j|}{|x_{i_0}|} \leq \sum_{j \neq i_0} |M_{i_0 j}| = R_{i_0}$.

That’s it for today

- Solutions for hwk1, hwk2 are on the course webpage.
- Midterm on Monday.
- Good luck!