Lecture 8: April 12, 2023

Lecturer: Ali Vakilian (notes based on notes from Madhur Tulsiani)

1 Applications of SVD: least squares approximation

We discuss another application of singular value decomposition (SVD) of matrices. Let $a_1, \ldots, a_n \in \mathbb{R}^d$ be points which we want to fit to a low-dimensional subspace. The goal is to find a subspace S of $\mathbb{R}^d$ of dimension at most k to minimize $\sum_{i=1}^n (\text{dist}(a_i, S))^2$, where $\text{dist}(a_i, S)$ denotes the distance of a_i from the closest point in S. We first prove the following.

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

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

Thus, the goal is to find a set of k orthonormal vectors $u_1, \ldots, u_k$ to maximize the quantity $\sum_{i=1}^n \sum_{j=1}^k \langle a_i, u_j \rangle^2$. Let $A \in \mathbb{R}^{n \times d}$ be a matrix with the i^{th} row equal to a_i^T. Then, we need to find orthonormal vectors $u_1, \ldots, u_k$ to maximize $\|Au_1\|_2^2 + \cdots + \|Au_k\|_2^2$. We will prove the following.

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 $u_1, \ldots, u_k$*

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

Thus, the optimal solution is to take $S = \text{Span}(v_1, \ldots, v_k)$. We prove the above by induction on k. For $k = 1$, we note that

$$\|Au_1\|_2^2 \;=\; \left\langle u_1, A^T A u_1 \right\rangle \;\leq\; \max_{v \in \mathbb{R}^d \setminus \{0\}} \mathcal{R}_{A^T A}(v) \;=\; \sigma_1^2 \;=\; \|Av_1\|_2^2 .$$

To prove the induction step for a given $k \leq r$, define

$$V_{k-1}^{\perp} = \left\{ v \in \mathbb{R}^d \mid \langle v, v_i \rangle = 0 \;\; \forall i \in [k-1] \right\} .$$

First prove the following claim.

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: We only provide a sketch of the proof here. Let $S = \text{Span}(\{u_1, \ldots, u_k\})$. Note that $\dim\left(V_{k-1}^{\perp}\right) = d - k + 1$ and $\dim(S) = k$. Thus,

$$\dim\left(V_{k-1}^{\perp} \cap S\right) \;\geq\; k + (d-k+1) - d \;=\; 1\,.$$

Hence, there exists $u'_k \in V_{k-1}^{\perp} \cap S$ with $\left\|u'_k\right\| = 1$. Completing this to an orthonormal basis of S gives orthonormal $u'_1, \ldots, u'_k$ with the first and second properties. We claim that this already implies the third property (why?). ■

Thus, we can assume without loss of generality that the given vectors $u_1, \ldots, u_k$ are such that $u_k \in V_{k-1}^{\perp}$. Hence,

$$\|Au_k\|_2^2 \;\leq\; \max_{\substack{v \in V_{k-1}^{\perp} \\ \|v\|=1}} \|Av\|_2^2 \;=\; \sigma_k^2 \;=\; \|Av_k\|_2^2\,.$$

Also, by the inductive hypothesis, we have that

$$\|Au_1\|_2^2 + \cdots + \|Au_{k-1}\|_2^2 \;\leq\; \|Av_1\|_2^2 + \cdots + \|Av_{k-1}\|_2^2\,,$$

which completes the proof. The above proof can also be used to prove that SVD gives the best rank k approximation to the matrix A in Frobenius norm.

2 Bounding the eigenvalues: Gershgorin Disc Theorem

We will now see a simple but extremely useful bound on the eigenvalues of a matrix, given by the Gershgorin disc theorem. Many useful variants of this bound can also be derived from the observation that for any invertible matrix S, the matrices $S^{-1}MS$ and M have the same eigenvalues (prove it!).

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

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

If λ is an eigenvalue of M, then

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

Proof: Let $x \in \mathbb{C}^n$ be an eigenvector corresponding to the eigenvalue λ. Let $i_0 = \mathrm{argmax}_{i \in [n]}\{|x_i|\}$. Since x is an eigenvector, we have

$$Mx = \lambda x \quad \Rightarrow \quad \forall i \in [n] \;\; \sum_{j=1}^{n} M_{ij} x_j \;=\; \lambda x_i\,.$$

In particular, we have that for $i = i_0$,

$$\sum_{j=1}^{n} M_{i_0 j} x_j \;=\; \lambda x_{i_0} \;\;\Rightarrow\;\; \sum_{j=1}^{n} M_{i_0 j} \frac{x_j}{x_{i_0}} \;=\; \lambda \;\;\Rightarrow\;\; \sum_{j \neq i_0} M_{i_0 j} \frac{x_j}{x_{i_0}} \;=\; \lambda - M_{i_0 i_0}\,.$$

Thus, we have

$$|\lambda - M_{i_0 i_0}| \;\leq\; \sum_{j \neq i_0} \left|M_{i_0 j}\right| \cdot \left|\frac{x_j}{x_{i_0}}\right| \;\leq\; \sum_{j \neq i_0} \left|M_{i_0 j}\right| \;=\; R_{i_0}\,.$$

■