Technische Universität München

Summary of the lecture MA4800

Foundations in Data Analysis

Instructors: Prof. Felix Krahmer and Dr. Anna Veselovska

Contents

1 Linear Algebra Review

- We work on $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$.
- $A^H = \overline{(A^T)}$.
- A Hermitian matrix A satisfies $A = A^H$.
- $A^{(i)}$ are rows and $A_{(j)}$ are the columns.
- $A^{(i)} = (a_{ij})_{j \in J}$ and $A_{(j)} = (a_{ij})_{i \in I} = \left(A^T\right)^{(j)}$
- The matrix-vector product between $A \in \mathbb{K}^{I \times J}$ and $x \in \mathbb{K}^I$ results in the vector in $Ax \in K'$ with entries

1.1 Matrices

$$(Ax)_i = \sum_{j \in J} a_{ij} x_j.$$

1.2 Matrix multiplication

The matrix-matrix product between $A \in \mathbb{K}^{I \times J}$ and $B \in \mathbb{K}^{J \times L}$ yields the matrix in $\mathbb{K}^{I \times L}$ with entries

$$(AB)_{i\ell} = \sum_{j \in J} A_{ij} B_{j\ell}.$$

2 The Singular Value Decomposition

2.1 The leading singular vector

2.2 Principal components

2.3 Further singular vectors

2.4 Best k-rank approximation

2.5 The power method

Lemma 2.1 *Let $x \in \mathbb{R}^d$ be a unit d-dimensional vector of components $x = (x_1, \ldots, x_d)$ with respect to the canonical basis and picked uniformly at random from the sphere $\{x : \|x\|_2 = 1\}$. The probability that $|x_1| \geq \alpha > 0$ is at least $1 - C\alpha\sqrt{d}$ for some absolute constant.*

Proof

We want the probability of y picked uniformly at random from

$$B^d(1) = \left\{y \in \mathbb{R}^d, ||y||_2 \leq 1\right\}$$

satisfies $|y_1| > \alpha$. In other words, we are looking for the fraction of $B^d(1)$ that satisfies $|y_1| > \alpha$. This corresponds to

$$V_\alpha := \mathrm{Vol}(B^d(1) \cap \{y : |y_1| \leq \alpha\})$$

$$= \int_{y \in B^d(1) \cap \{y : |y_1| \leq \alpha\}} 1 dy$$

$$= \int_{-\alpha}^{\alpha} \left(\int_{\mathbb{R}^{d-1}} 1_{y_2^2+...+y_d^2 \leq 1-y_1^2} \, dy_2...dy_d \right) dy_1$$

$$= \int_{-\alpha}^{\alpha} \mathrm{Vol} \left(B^{d-1} \left(\sqrt{1-y_1^2} \right) \right) dy_1$$

Replacing $\mathrm{Vol} \left(B^{d-1} \left(\sqrt{1-y_1^2} \right) \right)$ with $(\sqrt{1-y_1^2})^{d-1} \mathrm{Vol} \left(B^{d-1}(1) \right)$ since the volume the unit ball with a factor proportional to radius in the power of $d-1$.

$$= \int_{-\alpha}^{\alpha} (\sqrt{1-y_1^2})^{d-1} \mathrm{Vol} \left(B^{d-1}(1) \right) dy_1$$

$$= \mathrm{Vol} \left(B^{d-1}(1) \right) \int_{-\alpha}^{\alpha} (1-y_1^2)^{(d-1)/2} dy_1$$

In the integral part, $\int_{-\alpha}^{\alpha} (1-y_1^2)^{(d-1)/2} dy_1$, notice that $(1-y_1^2)^{(d-1)/2} < 1$ in the whole integration domain. Thus we can write

$$= \mathrm{Vol} \left(B^{d-1}(1) \right) \int_{-\alpha}^{\alpha} (1-y_1^2)^{(d-1)/2} dy_1$$

$$\leq \mathrm{Vol} \left(B^{d-1}(1) \right) \int_{-\alpha}^{\alpha} 1 dy_1$$

$$= 2\alpha \mathrm{Vol} \left(B^{d-1}(1) \right)$$

Recall that volume of unit ball in d dimensions is asymptotically

$$V_1 = \frac{1}{\sqrt{d\pi}} \left(\frac{2\pi e}{d} \right)^{d/2}$$

Hence the probability $p = \Pr(\alpha \leq |y_1|)$ we are interested in satisfies asymptotically

$$p = \frac{V_\alpha}{V_1} \propto \frac{2\alpha \frac{1}{\sqrt{(d-1)\pi}} \left(\frac{2\pi e}{d-1} \right)^{(d-1)/2}}{\frac{1}{\sqrt{d\pi}} \left(\frac{2\pi e}{d} \right)^{d/2}} = \frac{2\alpha \frac{1}{\sqrt{(d-1)\pi}} \left(\frac{2\pi e}{d-1} \right)^{(d-1)/2}}{\frac{1}{\sqrt{d\pi}} \left(\frac{2\pi e}{d} \right)^{(d-1)/2} \left(\frac{2\pi e}{d} \right)^{1/2}}$$

We simplify the last term

$$= 2\alpha * \left(\frac{d}{d-1} \right)^{1/2} * \left(\frac{d}{d-1} \right)^{(d-1)/2} * \left(\frac{d}{2\pi e} \right)^{1/2}$$

$$= 2\alpha * \left(\frac{d}{\sqrt{2\pi e(d-1)}} \right) * \left(\frac{d}{d-1} \right)^{(d-1)/2}$$

Since $\frac{d}{d-1} = 1 + \frac{1}{d-1}$

$$= 2\alpha * \left(\frac{d}{\sqrt{2\pi e(d-1)}} \right) * \left(1 + \frac{1}{d-1} \right)^{(d-1)/2}$$

We modify the power of the same term, to show it as

$$= 2\alpha * \left(\frac{d}{\sqrt{2\pi e(d-1)}} \right) * \left(\left(1 + \frac{1}{d-1} \right)^{(d-1)} \right)^{1/2}$$

Recall that

$$e = \lim_{n \to \infty} (1 + 1/n)^n$$

Thus this term is bounded with $\sqrt{e}$

$$\leq 2\alpha * \left(\frac{d}{\sqrt{2\pi e(d-1)}} \right) * \sqrt{e}$$

We reformulate as

$$= \alpha\sqrt{d}\sqrt{\frac{2d}{\pi(d-1)}}$$

Since $\sqrt{\frac{d}{d-1}} \leq 2$ for $d \geq 2$

$$\leq \frac{2\sqrt{2}}{\pi}\alpha\sqrt{d}$$

Given that all of this only holds asymptotically; we might need another multiplicative constant to make it hold in general. Hence the constant C in the theorem.

$$p \leq C\alpha\sqrt{d}$$

This bounds the probability $p = \Pr(\alpha \leq |y_1|) \leq C\alpha\sqrt{d}$. Considering the probability of the complement event the bounds $1 - \Pr(\alpha > |y_1|) \leq C\alpha\sqrt{d}$ can be stated as

$$1 - C\alpha\sqrt{d} \leq \Pr(\alpha > |y_1|).$$

Remark 2.1 *Notice that in the previous result essentially shows also that, independently of the dimension d, the* $x_1 = \langle x, u_1 \rangle$ *component of a random unit vector x with respect to any orthonormal basis* $\{u_1, ..., u_d\}$[1] *is bounded away from zero with overwhelming probability.*

Remark 2.2 *Consider the isometric mapping* $(a, b) \to a + bi$ *from* $\mathbb{R}^2$ *to* $\mathbb{C}$*. The previous result extends to random unit vectors in* $\mathbb{C}^d$ *simplify by modifying the statement as follows: The probability that, for a randomly chosen unit vector* $z \in \mathbb{C}^d$, $|z_1| \geq \alpha > 0$ *holds is at least* $1 - C\alpha\sqrt{2d} = 1 - C'\alpha\sqrt{d}$*.*

It is important to note that remark 2.1 and remark 2.2 holds with any orthonormal basis by rotating it to coincide with the canonical basis.

Theorem 2.2 *Let* $A \in \mathbb{K}^{I \times J}$ *and* $x \in \mathbb{K}^I$*. Let V be the space spanned by the left singular vectors of A corresponding to singular values greater than* $(1-\epsilon)\sigma_1$*. Let* $m \in \Omega\left(\frac{\ln(d/\epsilon)}{\epsilon}\right)$*. Let* w^* *be the unit vector after m iterations of the power method, namely,*

$$w^* = \frac{\left(AA^H\right)^m x}{|| \left(AA^H\right)^m x||_2} \tag{1}$$

The probability that w^* *has a component of at most l, where* $l \in O\left(\frac{\epsilon}{\alpha d}\right)$*, orthogonal to V is at least* $1 - C\alpha\sqrt{d}$ *i.e.* $1 - C\alpha\sqrt{d} < \Pr\left(\|Proj_{V^\perp}(w^*)\|_2 < l\right)$*.*

Proof

Let the SVD of A be given by

$$A = \sum_{k=1}^{r} \sigma_k u_k v_k^H$$

If the rank of A is less than $n = |I|$ we complete the orthonormal set of vectors $\{u_1, ..., u_r\}$ into a full orthogonal basis $\{u_1, ..., u_n\}$ of the n-dimensional space. We can expand x in the terms of this basis as

$$x = \sum_{k=1}^{n} \langle x, u_k \rangle u_k$$

We set $\sigma_k = 0$ for $k > r$ so that we can write A as

$$A = \sum_{k=1}^{n} \sigma_k u_k v_k^H$$

It follows that

$$(AA^H)^m x = \sum_{k=1}^{n} \sigma_k^{2m} u_k u_k^H x = \sum_{k=1}^{n} \sigma_k^{2m} u_k \langle x, u_k \rangle$$

By lemma 2.1, remark 2.1 and remark 2.2 one has $|\langle x_1, u_1 \rangle| \geq \alpha > 0$ with probability at least $1 - C\alpha\sqrt{d}$. We choose r_ϵ such that $\sigma_1, ..., \sigma_{r_\epsilon}$ are the singular values of A that are greater or equal to $(1-\epsilon)\sigma_1$ and $\sigma_{r_\epsilon+1}, ..., \sigma_n$ are those that are less than $(1-\epsilon)\sigma_1$. Notice that $V = \text{span}\{\sigma_1, ..., \sigma_{r_\epsilon}\}$ and $V^\perp = \text{span}\left\{\sigma_{r_{\epsilon+1}}, ..., \sigma_n\right\}$. The component of w^* orthogonal to $V^\perp$ is $\text{Proj}_{V^\perp}(w^*)$ which can be written as

$$\text{Proj}_{V^\perp}(w^*) = \frac{\text{Proj}_{V^\perp}\left(\left(AA^H\right)^m x\right)}{\|\left(AA^H\right)^m x\|_2} \tag{2}$$

We find denominator of equation 2 by Pythagoras-Fourier theorem

$$||(AA^H)^m x||_2^2 = \sum_{k=1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2 \tag{3}$$

$$\sum_{k=1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2 \geq \sigma_1^{4m} |\langle x, u_1 \rangle|^2 \geq \sigma_1^{4m} \alpha^2 \tag{4}$$

with probability at least $1 - C\alpha\sqrt{d}$. To find the nominator of equation 2, we check component of $(AA^H)^m x$ that is orthogonal to $V = span\{u_1, \ldots, u_{r_\epsilon}\}$, namely,

$$\text{Proj}_{V^\perp}\left(\left(AA^H\right)^m x\right) = \sum_{k=1}^{n} \sigma_k^{2m} |\langle x, u_k \rangle|_2 = \sum_{k=r_\epsilon+1}^{n} \sigma_k^{2m} |\langle x, u_k \rangle|_2 \tag{5}$$

$$\text{Proj}_{V^\perp}\left(\left(AA^H\right)^m x\right) \leq (1-\epsilon)^{2m} \sigma_1^{2m} \sum_{k=r_\epsilon+1}^{n} |\langle x, u_k \rangle|_2 \leq (1-\epsilon)^{2m} \sigma_1^{2m} \tag{6}$$

since $\sum_{k=r_\epsilon+1}^{n} \leq ||x||_2^2 = 1$ and $(1-\epsilon)\sigma_1 > \sigma_k$ for $r_\epsilon < k$.

By using 3 and 5 we find squared norm of the component of w^* orthogonal to V, that is $\|\text{Proj}_{V^\perp}(w^*)\|_2^2$, as

$$\|\text{Proj}_{V^\perp}(w^*)\|_2^2 = \frac{\sum_{k=r_\epsilon+1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2}{\sum_{k=1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2}$$

We bound this term by using the relations 4 and 6

$$\|\text{Proj}_{V^\perp}(w^*)\|_2^2 = \frac{\sum_{k=r_\epsilon+1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2}{\sum_{k=1}^{n} \sigma_k^{4m} |\langle x, u_k \rangle|^2} \leq \frac{(1-\epsilon)^{4m} \sigma_1^{4m}}{\alpha^2 \sigma_1^{4m}} = \frac{(1-\epsilon)^{4m}}{\alpha^2}$$

Thus, by taking the square root we have

$$\|\text{Proj}_{V^\perp}(w^*)\|_2 \leq \frac{(1-\epsilon)^{2m}}{\alpha}$$

In terms of *Big O* notation we have

$$\|\mathrm{Proj}_{V^\perp}(w^*)\|_2 \in \mathcal{O}\left(\frac{(1-\epsilon)^{2m}}{\alpha}\right)$$

Notice that $1-\epsilon$ is a linear approximation of $e^{-\epsilon}$. Similarly, $(1-\epsilon)^{2m}$ approximates e^{-2m} for small ϵ. Using this approximation,

$$\|\mathrm{Proj}_{V^\perp}(w^*)\|_2 \in \mathcal{O}\left(\frac{e^{-2\epsilon m}}{\alpha}\right)$$

Recall that $m \in \Omega\left(\frac{\ln(d/\epsilon)}{\epsilon}\right)$. This is another way of saying there exists m_0 and some constant $c > 0$ such that $m \geq c\frac{\ln(d/\epsilon)}{d}$ for all $m > m_0$. Similarly, this also means there exists m_0 and some constant $c > 0$ such that $-\frac{m}{c} \leq -\frac{\ln(d/\epsilon)}{d}$ for all $m > m_0$. Since exponentiation is a non-decreasing function $e^{-\frac{m}{c}} \leq e^{-\frac{\ln(d/\epsilon)}{\epsilon}} = e^{\frac{\ln(\epsilon/d)}{\epsilon}} = (\epsilon/d)^{1/\epsilon}$. We have

$$e^{-\frac{m}{c}} \leq (\epsilon/d)^{1/\epsilon}$$

for some constant $c > 0$ and $m > m_0$. We take the power of ϵ of both sides

$$e^{-\frac{m\epsilon}{c}} \leq \frac{\epsilon}{d}$$

Let $c_1 = c/2$

$$e^{-\frac{2m\epsilon}{c_1}} \leq \frac{\epsilon}{d}$$

For some constant $e^{-1/c_1} > 0$ and all $m > m_0$. We divide both sides with α

$$\frac{e^{-\frac{2m\epsilon}{c_1}}}{\alpha} \leq \frac{\epsilon}{\alpha d}$$

$$\alpha^{-1}e^{-\frac{2m\epsilon}{c_1}} \leq \frac{\epsilon}{\alpha d}$$

Which means

$$\frac{e^{-2m\epsilon}}{\alpha} \in \mathcal{O}\left(\frac{\epsilon}{\alpha d}\right)$$

Consequently

$$\|\mathrm{Proj}_{V^\perp}(w^*)\|_2 \in \mathcal{O}\left(\frac{\epsilon}{\alpha d}\right)$$