Advanced Linear Algebra Week 1 Day 1

2018/09/10 – Jonathan Hayase, updated by Prof. Weiqing Gu

1 Important things about eigenvalues and eigenvectors

Suppose we have $A\boldsymbol{x}_i = \lambda \boldsymbol{x}_i$ for $i = 1, 2, \ldots, n$. Then we have

$$(A\boldsymbol{x}_1, A\boldsymbol{x}_2, \ldots, A\boldsymbol{x}_n) = (\lambda_1 \boldsymbol{x}_1, \lambda_2 \boldsymbol{x}_2, \ldots, \lambda_n \boldsymbol{x}_n).$$

Vectorizing this yields:

$$A \underbrace{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}_{P} = \underbrace{(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)}_{P} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{D}$$

$$AP = PD \implies A = PDP^{-1}$$

This is useful because, when calculating high powers of A we get:

$$\begin{aligned} A^k &= \left(PDP^{-1}\right)^k && \text{for } k = 10,000 \\ &= \underbrace{PD\cancel{P^{-1}P}DP^{-1} \cdots PDP^{-1}}_{k \text{ times}} \\ &= PD^kP^{-1} \\ &= P \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{bmatrix} \end{aligned}$$

2 Least squares example

Suppose we have

$$y = a_0 + a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n$$

But really, we have

$$\mathrm{y}^{(N)} = \theta_0 + \theta_1 \boldsymbol{x}_1^{(N)} + \cdots + \theta_n \boldsymbol{x}_n^{(N)}$$

For $n \ll N$.

We can write this in vectorized form as:

$$\begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & x_1^{(1)} & \cdots & x_n^{(1)} \\ \vdots & & \ddots & \vdots \\ 1 & x_1^{(N)} & \cdots & x_n^{(N)} \end{bmatrix}}_{\text{design matrix}} \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_n \end{bmatrix}$$

So then we have:

$$y = X\boldsymbol{\theta}$$

$$X^T y = X^T X\boldsymbol{\theta} \implies \boldsymbol{\theta} = \left(X^T X\right)^{-1}\left(X^T X\right)$$

$$\begin{aligned}\left(X^T X\right)^T &= X^T\left(X^T\right)^T \\ &= X^T X\end{aligned}$$

So $X^T X$ is symmetric.

Symmetric matrices are diagonalizable.

Recall $A_{n\times n}$ Symmetric

1. All the eigenvalues of A are real
2. All eigenvectors belonging to distinct eigenvalues are orthogonal

Suppose we have

$$A = \left[\begin{smallmatrix}2&1&1\\1&2&1\\1&1&2\end{smallmatrix}\right]$$

Then we have

$$\begin{aligned}0 &= \det\left(A - \lambda I\right) \\ &= \begin{bmatrix}2-\lambda & 1 & 1\\ 1 & 2-\lambda & 1\\ 1 & 1 & 2-\lambda\end{bmatrix} \\ &= \begin{bmatrix}4-\lambda & 4-\lambda & 4-\lambda\\ 1 & 2-\lambda & 1\\ 1 & 1 & 2-\lambda\end{bmatrix} \\ &= (4-\lambda)\begin{bmatrix}1 & 1 & 1\\ 1 & 2-\lambda & 1\\ 1 & 1 & 2-\lambda\end{bmatrix} \\ &= (4-\lambda)\begin{bmatrix}1 & 1 & 1\\ 0 & 1-\lambda & 0\\ 0 & 0 & 1-\lambda\end{bmatrix} \\ &=\end{aligned}$$

Suppose we have a bunch of vectors, which are not orthonormal.

$$Av_1 = \lambda v_1$$

First, we normalize v_1 to get $u_1 = \frac{v_1}{\|v_1\|}$.

$$A\frac{v_1}{\|v_1\|} = \lambda\frac{v_1}{\|v_1\|}$$

$$v_2 \cdot u_1 = \|v_2\|\cancelto{1}{\|u_1\|}\cos\theta$$

Gram-Schmidt leads to the QR facotrization.

3 Orthonormal Eigenbases

Suppose we have an orthonormal eigenbasis

Then $P = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n)$ and we have $P^T P = I$, so $P^{-1} = P^T$ and

$$P^{-1}AP = D \implies P^T AP = D \implies A = PDP^T$$

If A is positive-definite then A has all positive eigenvalues

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$
$$A = P\sqrt{D}\sqrt{D}P^T = \left(P\sqrt{D}P^T\right)\left(P\sqrt{D}P^T\right) = LL = L^2$$

jburkert@headlandstech.com careers@headlandstech.com [1]

4 Inner Products

An inner-product is a positive bilinear form.

$$\mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

$$(v, w) \mapsto v \cdot w \text{ or } \langle v, w \rangle$$

Or if $f, g \in C[a, b]$ then

$$\langle f, g \rangle = \int_a^b fg \; \mathrm{d}x$$

Inner Products have the following properties

1. Bilinear

2. Symmetric

3. Positive Definite

Inner Products also induce a norm like so: Let $v = w$ then $\langle v, v \rangle = \|v\|^2$ so $\|v\| = \sqrt{\langle v, v \rangle}$. [2]

[1] Something something pfaffian

[2] Something something banach space

5 Vector Spaces of Infinite Dimension

Suppose you are working in the field $\mathbb{C}$.

Notation: ${}^{V}\!/\mathbb{C}$ is a vector space over $\mathbb{C}$

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}, \boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

Then we have $\langle z, w \rangle = \boldsymbol{z}_1\overline{\boldsymbol{w}_1} + \cdots + \boldsymbol{z}_n\overline{\boldsymbol{w}_n}$

Now suppose we have ${}^{V}\!/\mathbb{C}$ of $n = \infty$ dimensions. We see that $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = \|z_1\|^2 + \cdots + \|z_n\|^2 \geq 0$. And we have $\|z\| = 0$ iff $\|z_k\| = 0$ for $k \in \{1, 2 \ldots, k\}$.

Also we have:

$$\overline{\langle \boldsymbol{z}, \boldsymbol{w} \rangle} = \overline{\boldsymbol{z}_1\overline{\boldsymbol{w}_1} + \cdots + \boldsymbol{z}_n\overline{\boldsymbol{w}_n}} = \overline{\boldsymbol{z}_1}\boldsymbol{w}_1 + \cdots + \overline{\boldsymbol{z}_n}\boldsymbol{w}_n$$

$$\langle c\boldsymbol{z}_1 + \boldsymbol{z}_2, \boldsymbol{w} \rangle = c\langle \boldsymbol{z}_1, \boldsymbol{w} \rangle + \langle \boldsymbol{z}_2, \boldsymbol{w} \rangle$$

Definition:

Let X be a vector space over field K.

A scalar (or inner) product on X is

$$\langle \cdot, \cdot \rangle : X \times X \to K$$

Satisfying For any $k \in K$ and $x, y \in X$

1. $\langle x, x \rangle \geq 0$
2.

6 Examples

$$\ell^2(\mathbb{R}) = \left\{ \{x_i\}_{i=0}^{\infty} \mid x_i \in \mathbb{R} \wedge \sum x_i^2 \leq \infty \right\}$$

$$\langle \{x_i\}_{i=0}^{\infty}, \{y_i\}_{i=0}^{\infty} \rangle \stackrel{?}{=} \sum_{i=0}^{\infty} x_i y_i$$

Does this converge. (Yes, but we will see this later.)

Now consider $L^2[a, b] = \left\{ f \mid f : [a, b] \to \mathbb{R} \wedge \int_a^b \left(f(x)\right)^2 \, \mathrm{d}x \text{ exists and is finite} \right\}$

$$\langle f, g \rangle = \int_a^b fg \, \mathrm{d}x$$

Now consider $\ell^2(\mathbb{C}) \left\{ \{z_i\}_{i=0}^{\infty} \mid z_i \in \mathbb{C} \wedge \sum \|z_i\| < \infty \right\}$ And we define

$$\langle \{z_i\}_{i=0}^{\infty}, \{w_i\}_{i=0}^{\infty} \rangle = \sum_{i=0}^{\infty} z_i \overline{w_i}$$

This also converges.

7 Zorn's Lemma

Recall

Totally ordered: is a relation R (representing $\leq$) on a set A.

1. Antisymmetry: for all $a, b \in A$ if $a \leq b \wedge b \leq a \implies a = b$.

2. Transitivity: For all $a, b, c \in A$ if $a \leq b \wedge b \leq c \implies a \leq c$

3. $a \leq b$ or $b \leq a$

(You may substitue $x \leq y$ with $R(x, y)$ in the above.)

Partially ordered: Example: Let a set $A = \{a, b, c\}$ now, consider the powerset $\mathcal{P}(A) = \big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, b\}, \{a$
Inclusion "$\subseteq$" is a partial order on $\mathcal{P}(A)$, but not a total order.

You can make a Hasse Diagram representing this ordering

$$\begin{tikzcd}
 & & A & & \\
\{a,b\} \arrow[urr] & & \{a,c\} \arrow[u] & & \{b,c\} \arrow[ull] \\
\{a\} \arrow[u] \arrow[urr] & & \{b\} \arrow[ull] \arrow[urr] & & \{c\} \arrow[ull] \arrow[u] \\
 & & \emptyset \arrow[ull] \arrow[u] \arrow[urr] & &
\end{tikzcd}$$

Draw an arrow from x to y if $x \,\mathrm{R}\, y$ (i.e. $x \subseteq y$). We call a sequence $\emptyset \subseteq \{b\} \subseteq \{b, c\} \subseteq A$ a chain.

Note: we can't compare $\{b\}$ and $\{c\}$.

Zorn's Lemma: Let P be a partially ordered set such that every chain has an upper bound in P. Then the set P contains at least one maximum element.

Idea of proof:

Recall In Linear Algebra you will use three kinds of basic proving techniques:

1. Induction

2. Contradiction

   Example: (Zorn's Lemma) Use Hasse diagram of (P, R). Suppose there were no such maximal element. Then a chain can go forever. There exists some chain in the Hasse diagram which has larger and larger elements, without bound. But then this contradicts the assertion that all chains are upper bounded.

3. Build a bridge[a]

   Example: Prove "inverse of $A_{n\times n}$ is unique." Suppose $AB = BA = I$ and $AC = CA = I$ want to show $B = C$. Then we build the following bridge:

$$B = BI = B(AC) = (BA)C = IC = C$$

[a]You basically start from the definition (and maybe also your goal) and meet in the middle.

8 Dual Space

8.1 Motivation

Theorem: Let $x_i \in \mathbb{R}^n$ be a sparse vector. Let A be a $m \times n$ random matrix with $A_{ij} \sim \mathcal{N}(0, {}^1/_m)$. Then x_0 is the unique optimal solution minimizing $\|x\|_1$ subject to $Ax = Ax_0$.

Changing point view: Let A change, x_0 is fixed.

$$m \geq (1+\epsilon)2S\log n + s + 1$$

With probability at least

$$1 - 2n^{-\left[\sqrt{1+{}^\epsilon/_{2s}+\epsilon} - \sqrt{1+{}^\epsilon/_{2s}}\right]^2}$$

8.2 Definition

Recall If we have an orthonormal basis $\{u_1, \ldots, u_n\}$ of a vector space ${}^V/_\mathbb{R}$ if we have any vector $\boldsymbol{v} = a_1\boldsymbol{x}_n + \cdots + a_n\boldsymbol{u}_n$.

$$a_1 = \boldsymbol{v}_1 \cdot \boldsymbol{u}_1$$
$$\vdots$$
$$a_n = \boldsymbol{v}_n \cdot \boldsymbol{u}_n$$

8.3 Linear Functional

Let V be a vector space over a field $\mathbb{F}$. Then a linear maps $\ell : V \to \mathbb{F}$ is also called a linear functional on V.

1. Basic example: Define $L : \mathbb{F}^n \to \mathbb{F}$ where

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} \mapsto a_1\boldsymbol{x}_1 + \cdots + a_n\boldsymbol{x}_n = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_n \end{bmatrix} = \ell(\boldsymbol{x})$$

   If you just multiply by a row vector, it's called functional.

2. $\operatorname{tr} A =$

$$\mathbb{F}^n \to \mathbb{F}$$
$$\operatorname{tr}(cA + B) = c \operatorname{tr} A + \operatorname{tr} B$$

3. "Evaluation at a point" is a linear function on $P_n(\mathbb{C})$

$$L : P_n(\mathbb{C}) \to \mathbb{C}$$

4. "Integration"

$$C\,[a, b] \to \mathbb{R}$$
$$L(f) \to \int_a^b f(t)\ \mathrm{d}t$$

5. "Derivative"

$$L : C^2 \to \mathbb{R}$$