Mathematics of Data Sceience

Tutorial, October 20

Hung-Hsu Chou

Mathematics of Information Processing, RWTH

Winter Semester 2021/2022

Outline

Outline

Quote

The supreme task of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them.

– Albert Einstein

Outline

Prediction

Question

Recall our potential teammates x, y, z.

	Intelligence	*Friendliness*	*Punctuality (On Time)*
x	9	5	5
y	6	8	6
z	7	7	7

Suppose we have worked with y and z already, and the experience is bad with y and good with z. What kind of experience would we expect with x?

What are the reasons for those experience?

What are the reasons for those experience?

One way to derive the right mapping for score, so that we can compare. Formally, we assume that $f(y) < f(z)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the underlying mapping, and would like to find $f(x)$. What are the difficulties here?

What are the reasons for those experience?

One way to derive the right mapping for score, so that we can compare. Formally, we assume that $f(y) < f(z)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the underlying mapping, and would like to find $f(x)$. What are the difficulties here?

- Uniqueness: there are infinitely many f that satisfy $f(y) > f(z)$.

What are the reasons for those experience?

One way to derive the right mapping for score, so that we can compare. Formally, we assume that $f(y) < f(z)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the underlying mapping, and would like to find $f(x)$. What are the difficulties here?

- Uniqueness: there are infinitely many f that satisfy $f(y) > f(z)$. Possible solution: impose constraints on f, collect more data,...etc.

What are the reasons for those experience?

One way to derive the right mapping for score, so that we can compare. Formally, we assume that $f(y) < f(z)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the underlying mapping, and would like to find $f(x)$. What are the difficulties here?

- Uniqueness: there are infinitely many f that satisfy $f(y) > f(z)$. Possible solution: impose constraints on f, collect more data,...etc.
- Precision: good and bad are rough feelings that may not be precisely transformed into numbers.

What are the reasons for those experience?

One way to derive the right mapping for score, so that we can compare. Formally, we assume that $f(y) < f(z)$, where $f : \mathbb{R}^3 \to \mathbb{R}$ is the underlying mapping, and would like to find $f(x)$. What are the difficulties here?

- Uniqueness: there are infinitely many f that satisfy $f(y) > f(z)$. Possible solution: impose constraints on f, collect more data,...etc.
- Precision: good and bad are rough feelings that may not be precisely transformed into numbers. Possible solution: reduce the range of f.

Did we have similar experience?

Did we have similar experience?

Another way to compare the new person with the old ones. If x is more similar to y, then it is more likely to be a bad experience. If x is more similar to z, then it is more likely to be a good experience.

Did we have similar experience?

Another way to compare the new person with the old ones. If x is more similar to y, then it is more likely to be a bad experience. If x is more similar to z, then it is more likely to be a good experience.

The convention way to compute similarity is inner product. If we define $\langle x, z\rangle = \sum_{i=1}^{3} x_i z_i$, then

$$\langle x, y\rangle = 124, \quad \langle x, z\rangle = 133.$$

Since $\langle x, y\rangle < \langle x, z\rangle$, x is more similar to z, and hence possibly good experience.

Outline

Symmetric Inner Product

Given a $\mathbb{R}$-vector space V, a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called a *symmetric inner product* if the following properties hold:

Symmetric Inner Product

Given a $\mathbb{R}$-vector space V, a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is called a *symmetric inner product* if the following properties hold:

Positive Definiteness For all $v \in V$, $\langle v, v \rangle \geq 0$.

Definiteness If $\langle v, v \rangle = 0$ if and only if $v = 0$.

Linearity For all $\lambda \in \mathbb{R}$ and $u, v, w \in V$,
$\langle v + \lambda u, w \rangle = \langle v, w \rangle + \lambda \langle u, w \rangle$.

Symmetry For all $u, v \in V$, $\langle u, v \rangle = \langle v, u \rangle$.

Norm

Given a $\mathbb{R}$-vector space V, a mapping $\|\cdot\|: V \to \mathbb{R}$ is called a *norm* if the following properties hold:

Norm

Given a $\mathbb{R}$-vector space V, a mapping $\|\cdot\| : V \to \mathbb{R}$ is called a *norm* if the following properties hold:

Definiteness If $\|v\| = 0$, if and only if $v = 0$.

Linearity For all $\lambda \in \mathbb{R}$ and $v \in V$, $\|\lambda v\| = |\lambda|\|v\|$.

Triangle Inequality For all $u, v \in V$, $\|u + v\| \leq \|u\| + \|v\|$.

Outline

Exercise 1a

Prove that the mapping $\| \cdot \| : V \to \mathbb{R}$ with

$$\|v\| := \sqrt{\langle v, v \rangle}$$

is a norm on V if $\langle \cdot, \cdot \rangle$ is an symmetric inner product on V.

Exercise 1a

For notation simplicity we will denote $f(v) = \sqrt{\langle v, v \rangle}$. Our goal is to prove $f(v)$ satisfies Definiteness, Linearity, Triangle Inequality,

assuming that

$\langle v, v \rangle$ satisfies Positive Definiteness, Definiteness, Linearity, Symmetry.

Solution to 1a

Recall Definiteness: we need to prove $f(v) = 0$ if and only if $v = 0$.

Solution to 1a

Recall Definiteness: we need to prove $f(v) = 0$ if and only if $v = 0$.

Proof.

$$
\begin{aligned}
f(v) = 0 &\iff f(v)^2 = \langle v, v \rangle = 0 \\
&\iff v = 0. && (\textit{Definiteness})
\end{aligned}
$$

Solution to 1a

Recall Linearity: we need to prove that for all $\lambda \in \mathbb{R}$ and $v \in V$, $f(\lambda v) = |\lambda| f(v)$. Let $v \in V$ and $\lambda \in \mathbb{R}$. We will first look at $f(\lambda v)^2$.

Solution to 1a

Recall Linearity: we need to prove that for all $\lambda \in \mathbb{R}$ and $v \in V$, $f(\lambda v) = |\lambda| f(v)$. Let $v \in V$ and $\lambda \in \mathbb{R}$. We will first look at $f(\lambda v)^2$.

Proof.

$$\begin{aligned} f(\lambda v)^2 &= \langle \lambda v, \lambda v \rangle = \lambda \langle v, \lambda v \rangle && (\textit{Linearity}) \\ &= \lambda \langle \lambda v, v \rangle && (\textit{Symmetry}) \\ &= \lambda^2 \langle v, v \rangle. && (\textit{Linearity}) \end{aligned}$$

Solution to 1a

Recall Linearity: we need to prove that for all $\lambda \in \mathbb{R}$ and $v \in V$, $f(\lambda v) = |\lambda| \, f(v)$. Let $v \in V$ and $\lambda \in \mathbb{R}$. We will first look at $f(\lambda v)^2$.

Proof.

$$\begin{aligned} f(\lambda v)^2 &= \langle \lambda v, \lambda v \rangle = \lambda \langle v, \lambda v \rangle && (\textit{Linearity}) \\ &= \lambda \langle \lambda v, v \rangle && (\textit{Symmetry}) \\ &= \lambda^2 \langle v, v \rangle. && (\textit{Linearity}) \end{aligned}$$

Hence

$$f(\lambda v) = \pm |\lambda f(v)| = |\lambda| \, f(v) \qquad (\textit{PositiveDefiniteness}).$$

Solution to 1a

Recall Triangle Inequality: we need to prove that for all $u, v \in V$, $f(u+v) \leq f(u) + f(v)$. Let $u, v \in V$ and consider $f(u+v)^2$.

Solution to 1a

Recall Triangle Inequality: we need to prove that for all $u, v \in V$, $f(u+v) \leq f(u) + f(v)$. Let $u, v \in V$ and consider $f(u+v)^2$.

Proof.

$$
\begin{aligned}
f(u+v)^2 &= \langle u+v, u+v \rangle \\
&= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle && (\textit{Linearity}) \\
&= \langle u, u \rangle + 2\langle u, v \rangle + \langle v, v \rangle && (\textit{Symmetry}) \\
&\leq f(u)^2 + 2f(u)f(v) + f(v)^2. && (\text{Cauchy Schwarz}) \\
&= (f(u) + f(v))^2
\end{aligned}
$$

Since f has Definiteness, $f(u+v) \leq f(u) + f(v)$.

Exercise 1b

Prove that The Frobenius scalar product defined by

$$\langle A, B\rangle_F := \operatorname{tr}\left(AB^T\right)$$

is a symmetric inner product on $\mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n}$ and the associated norm is the Frobenius norm

$$\|A\|_F = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{\frac{1}{2}}.$$

Exercise 1b

Our goal is to prove that $\langle \cdot, \cdot \rangle_F$ satisfies Positive Definiteness, Definiteness, Linearity, Symmetry.

Exercise 1b

Our goal is to prove that $\langle \cdot, \cdot \rangle_F$ satisfies Positive Definiteness, Definiteness, Linearity, Symmetry. Let us first prove a simple relation: for $A, B \in \mathbb{R}^{m\times n}$,

$$\begin{aligned} \langle A, B\rangle_F = tr(AB^T) = \sum_{i=1}^{m} (AB^T)_{ii} \\ = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B^T_{ji} \\ = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}. \end{aligned} \tag{1}$$

In the following proofs we will let $A, B, C \in \mathbb{R}^{m\times n}$ be a matrix.

Solution 1b

Recall Positive Definiteness, we need to prove that for all $A \in \mathbb{R}^{m \times n}$, $\langle A, A \rangle_F \geq 0$.

Solution 1b

Recall Positive Definiteness, we need to prove that for all $A \in \mathbb{R}^{m \times n}$, $\langle A, A \rangle_F \geq 0$.

Proof. By (1),

$$\langle A, A \rangle_F = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2 \geq 0. \tag{2}$$

Solution 1b

Recall Positive Definiteness, we need to prove that $\langle A, A \rangle_F = 0$, if and only if $A = 0$.

Solution 1b

Recall Positive Definiteness, we need to prove that $\langle A, A\rangle_F = 0$, if and only if $A = 0$.

Proof. By (2),

$$\begin{aligned}\langle A, A\rangle_F = 0 &\iff \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2 = 0\\ &\iff A_{ij} = 0 \quad \forall\, i, j.\end{aligned}$$

Solution 1b

Recall Linearity, we need to prove that for all $\lambda \in \mathbb{R}$ and $A, B, C \in \mathbb{R}^{m\times n}$, $\langle A + \lambda B, C\rangle_F = \langle A, C\rangle_F + \lambda\langle B, C\rangle_F$.

Solution 1b

Recall Linearity, we need to prove that for all $\lambda \in \mathbb{R}$ and $A, B, C \in \mathbb{R}^{m \times n}$, $\langle A + \lambda B, C \rangle_F = \langle A, C \rangle_F + \lambda \langle B, C \rangle_F$.

Proof. By (1),

$$\begin{aligned}\langle A + \lambda B, C \rangle_F &= \sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} + \lambda B_{ij}) C_{ij} \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} C_{ij} + \lambda \sum_{i=1}^{m} \sum_{j=1}^{n} B_{ij} C_{ij} \\ &= \langle A, C \rangle_F + \lambda \langle B, C \rangle_F.\end{aligned}$$

Solution 1b

Recall Symmetry, we need to prove that for all $A, B \in \mathbb{R}^{m \times n}$, $\langle A, B \rangle_F = \langle B, A \rangle$.

Solution 1b

Recall Symmetry, we need to prove that for all $A, B \in \mathbb{R}^{m\times n}$, $\langle A, B\rangle_F = \langle B, A\rangle$.

Proof. By (1),

$$\begin{aligned}\langle A, B\rangle_F &= \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}B_{ij}\\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} B_{ij}A_{ij} = \langle B, A\rangle_F.\end{aligned}$$

Exercise 2

Suppose $A \in \mathbb{R}^{m\times n}$, $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $Y = (\mathbb{R}^m, \|\cdot\|_Y)$. Prove that

$$\|A\|_{X\to Y} := \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X}$$

is a norm on $\mathbb{R}^{m\times n}$,

Solution 2a

Prove Definiteness.

Proof.

- If $\|A\|_{X\to Y} = 0$, then

$$\sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X} = 0 \; .$$

This implies that $\|Ax\|_Y = 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Solution 2a

Prove Definiteness.

Proof.

- If $\|A\|_{X\to Y} = 0$, then

$$\sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X} = 0\ .$$

  This implies that $\|Ax\|_Y = 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.
- Since $\|\cdot\|_Y$ is a norm, we find $Ax = 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Solution 2a

Prove Definiteness.

Proof.

- If $\|A\|_{X\to Y} = 0$, then

$$\sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_Y}{\|x\|_X} = 0\ .$$

  This implies that $\|Ax\|_Y = 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.
- Since $\|\cdot\|_Y$ is a norm, we find $Ax = 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.
- Therefore $\ker(A) = \mathbb{R}^n$ and we have $A = 0$.

Prove Linearity.

Proof.

- Since $\|\cdot\|_Y$ is a norm we have $\|\lambda y\|_Y = |\lambda|\|y\|_Y$ for all $y \in Y$.

Solution 2a

Prove Linearity.

Proof.

- Since $\| \cdot \|_Y$ is a norm we have $\|\lambda y\|_Y = |\lambda| \|y\|_Y$ for all $y \in Y$.
- For each $x \in X$ we have $Ax \in Y$.

Solution 2a

Prove Linearity.

Proof.

- Since $\|\cdot\|_Y$ is a norm we have $\|\lambda y\|_Y = |\lambda|\|y\|_Y$ for all $y \in Y$.
- For each $x \in X$ we have $Ax \in Y$.
- Hence

$$\begin{aligned}\|\lambda A\|_{X\to Y} &= \sup_{x\in X\setminus\{0\}} \frac{\|\lambda Ax\|_Y}{\|x\|_X} \\ &= \sup_{x\in X\setminus\{0\}} \frac{|\lambda|\|Ax\|_Y}{\|x\|_X} = |\lambda|\|A\|_{X\to Y} \,.\end{aligned}$$

Solution 2a

Prove Triangle Inequality.

Proof.

- The image of x under $A + B$ is $Ax + Bx$ and therefore $(A + B)x \in Y$.

Solution 2a

Prove Triangle Inequality.

Proof.

- The image of x under $A+B$ is $Ax+Bx$ and therefore $(A+B)x \in Y$.
- Again, since $\|\cdot\|_Y$ is a norm we find,

$$\|(A+B)x\|_Y = \|Ax+Bx\|_Y \leq \|Ax\|_Y + \|Bx\|_Y \ .$$

Solution 2a

Prove Triangle Inequality.

Proof.

- The image of x under $A+B$ is $Ax+Bx$ and therefore $(A+B)x \in Y$.
- Again, since $\|\cdot\|_Y$ is a norm we find,

$$\|(A+B)x\|_Y = \|Ax+Bx\|_Y \leq \|Ax\|_Y + \|Bx\|_Y \ .$$

- Again, plugging this into the definition of $\|A\|_{X\to Y}$ the claim follows.

Exercise 2b

Given $A \in \mathbb{R}^{m \times n}$, $X = (\mathbb{R}^n, \|\cdot\|_X)$ and $Y = (\mathbb{R}^m, \|\cdot\|_Y)$. Prove that

$$\|A\|_{\ell^2 \to \ell^2} = \max_{j=1,\dots,n} \sqrt{\lambda_j\left(A^T A\right)}$$

where $\lambda_j\left(A^T A\right)$ is the j-th eigenvalue of $A^T A$.

Solution 2b

Proof.

- By definition, $\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A^T Ax, x \rangle$.

Solution 2b

Proof.

- By definition, $\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A^T Ax, x \rangle$.
- Since $A^T A$ is symmetric positive definite, it can be expressed as $A^T A = UDU^T$ for some orthogonal matrix U and diagonal matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ with non-negative diagonal.

Solution 2b

Proof.

- By definition, $\|Ax\|_2^2 = \langle Ax, Ax\rangle = \langle A^T Ax, x\rangle$.
- Since $A^T A$ is symmetric positive definite, it can be expressed as $A^T A = UDU^T$ for some orthogonal matrix U and diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with non-negative diagonal.
- Hence $\|Ax\|_2^2 = \langle UDU^T x, x\rangle = \langle \sqrt{D}U^T x, \sqrt{D}U^T x\rangle = \|\sqrt{D}\tilde{x}\|_2^2$, where $\|\tilde{x}\|_2^2 = \|U^T x\|_2^2 = \langle UU^T x, x\rangle = \langle UU^T x, x\rangle = \|x\|_2^2$.

Proof.

- By definition, $\|Ax\|_2^2 = \langle Ax, Ax\rangle = \langle A^T Ax, x\rangle$.
- Since $A^T A$ is symmetric positive definite, it can be expressed as $A^T A = UDU^T$ for some orthogonal matrix U and diagonal matrix $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with non-negative diagonal.
- Hence $\|Ax\|_2^2 = \langle UDU^T x, x\rangle = \langle \sqrt{D}U^T x, \sqrt{D}U^T x\rangle = \|\sqrt{D}\tilde{x}\|_2^2$, where $\|\tilde{x}\|_2^2 = \|U^T x\|_2^2 = \langle UU^T x, x\rangle = \langle UU^T x, x\rangle = \|x\|_2^2$.
- By definition,

$$\|\sqrt{D}\tilde{x}\|_2^2 = \sum_{i=1}^n \lambda_i x_i^2 \leq \max_{j\leq n} \lambda_j \sum_{i=1}^n x_i^2 = \max_{j\leq n} \lambda_j \|\tilde{x}\|_2^2$$

Solution 2b

continuous. In summary we have that

$$
\begin{aligned}
\|A\|_{\ell^2\to\ell^2} &= \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_2}{\|x\|_2} \\
&= \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|\sqrt{D}\tilde{x}\|_2}{\|x\|_2} \\
&\leq \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\max_{j\leq n}\sqrt{\lambda_j}\|\tilde{x}\|_2}{\|\tilde{x}\|_2} \\
&= \max_{j\leq n}\sqrt{\lambda_j}
\end{aligned}
$$

Solution 2b

We have shown $\|A\|_{\ell^2\to\ell^2} \leq \max_{j\leq n}\sqrt{\lambda_j}$. It remains to show that $\|A\|_{\ell^2\to\ell^2} \geq \max_{j\leq n}\sqrt{\lambda_j}$.

Solution 2b

We have shown $\|A\|_{\ell^2 \to \ell^2} \leq \max_{j \leq n} \sqrt{\lambda_j}$. It remains to show that $\|A\|_{\ell^2 \to \ell^2} \geq \max_{j \leq n} \sqrt{\lambda_j}$.

Proof. Assume that λ_* is the largest eigenvalue of $A^T A$ and let u_* denote the corresponding eigenvector with $\|u_*\|_2 = 1$. Then

$$\|Au_*\|_2 = \sqrt{\langle A^T A u_*, u_* \rangle} = \sqrt{\langle \lambda_* u_*, u_* \rangle} = \sqrt{\lambda_*}.$$

Solution 2b

We have shown $\|A\|_{\ell^2\to\ell^2} \le \max_{j\le n}\sqrt{\lambda_j}$. It remains to show that $\|A\|_{\ell^2\to\ell^2} \ge \max_{j\le n}\sqrt{\lambda_j}$.

Proof. Assume that λ_* is the largest eigenvalue of A^TA and let u_* denote the corresponding eigenvector with $\|u_*\|_2 = 1$. Then

$$\|Au_*\|_2 = \sqrt{\langle A^TAu_*, u_*\rangle} = \sqrt{\langle \lambda_* u_*, u_*\rangle} = \sqrt{\lambda_*}.$$

Since

$$\sqrt{\lambda_*} \le \|A\|_{\ell^2\to\ell^2} \le \sqrt{\lambda_*},$$

we have that $\|A\|_{\ell^2\to\ell^2} = \sqrt{\lambda_*}$.

Exercise 3

Given the matrix $A \in \mathbb{R}^{3\times 3}$ where

$$A = \begin{pmatrix} 7 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 7 \end{pmatrix}$$

calculate a diagonal matrix $D \in \mathbb{R}^{3\times 3}$ and an orthogonal matrix $V \in \mathbb{R}^{3\times 3}$ such that $A = VDV^\top$.

Solution 3

Calculate the eigenvalues:

$$
\begin{aligned}
0 = \det(A - \lambda \mathrm{Id}) &= \det \begin{pmatrix} 7-\lambda & 0 & 4 \\ 0 & 4-\lambda & 0 \\ 4 & 0 & 7-\lambda \end{pmatrix} \\
&= (7-\lambda)^2(4-\lambda) - 4^2(4-\lambda) \\
&= (4-\lambda)[(7-\lambda)^2 - 4^2]
\end{aligned}
$$

Solution 3

Calculate the eigenvalues:

$$
\begin{aligned}
0 = \det(A - \lambda \mathrm{Id}) &= \det \begin{pmatrix} 7-\lambda & 0 & 4 \\ 0 & 4-\lambda & 0 \\ 4 & 0 & 7-\lambda \end{pmatrix} \\
&= (7-\lambda)^2(4-\lambda) - 4^2(4-\lambda) \\
&= (4-\lambda)[(7-\lambda)^2 - 4^2]
\end{aligned}
$$

The expression $(4-\lambda)[(5-\lambda)^2 - 16]$ is 0, if either of the terms $(4-\lambda)$ or $[(7-\lambda)^2 - 4^2] = (\lambda - 3)(\lambda - 11)$ is zero. Hence,

$$\lambda_1 = 11, \qquad \lambda_2 = 4 \qquad \text{and} \qquad \lambda_3 = 3\,.$$

Solution 3

λ_1.

$$\ker(A - 11 \cdot Id) = \ker \begin{pmatrix} -4 & 0 & 4 \\ 0 & -7 & 0 \\ 4 & 0 & -4 \end{pmatrix} = \text{span}\{(1,0,1)^T\}$$

Solution 3

λ_1.

$$\ker(A - 11 \cdot Id) = \ker \begin{pmatrix} -4 & 0 & 4 \\ 0 & -7 & 0 \\ 4 & 0 & -4 \end{pmatrix} = \operatorname{span}\{(1,0,1)^T\}$$

λ_2.

$$\ker(A - 4 \cdot Id) = \ker \begin{pmatrix} 3 & 0 & 4 \\ 0 & 0 & 0 \\ 4 & 0 & 3 \end{pmatrix} = \operatorname{span}\{(0,1,0)^T\}$$

Solution 3

λ_1.

$$\ker(A - 11 \cdot Id) = \ker \begin{pmatrix} -4 & 0 & 4 \\ 0 & -7 & 0 \\ 4 & 0 & -4 \end{pmatrix} = \text{span}\{(1, 0, 1)^T\}$$

λ_2.

$$\ker(A - 4 \cdot Id) = \ker \begin{pmatrix} 3 & 0 & 4 \\ 0 & 0 & 0 \\ 4 & 0 & 3 \end{pmatrix} = \text{span}\{(0, 1, 0)^T\}$$

λ_2.

$$\ker(A - 3 \cdot Id) = \ker \begin{pmatrix} 4 & 0 & 4 \\ 0 & 1 & 0 \\ 4 & 0 & 4 \end{pmatrix} = \text{span}\{(1, 0, -1)^T\}$$

Solution 3

After normalizing the vectors, we find that the vectors

$$v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad v_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

form an orthonormal basis of $\mathbb{R}^3$.

Solution 3

After normalizing the vectors, we find that the vectors

$$v_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}, \qquad v_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}, \qquad v_3 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\end{pmatrix}$$

form an orthonormal basis of $\mathbb{R}^3$. Hence we have found

$$A = VDV^T \qquad \text{with} \qquad D = \begin{pmatrix}11 & 0 & 0\\0 & 4 & 0\\0 & 0 & 3\end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 0 & 1\\0 & \sqrt{2} & 0\\1 & 0 & -1\end{pmatrix}.$$

Exercise 4

Let $W \subset \mathbb{R}^m$ be a linear subspace of dimension k with orthonormal basis $w_1, \ldots, w_k \in \mathbb{R}^m$ and let $u \in \mathbb{R}^m$. Prove the following statements:

4a) The minimizer $\hat{w}$ of $\min_{w \in W} \|u - w\|_2$ exists, is unique and is given by $\hat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$.

4b) The difference vector $u - \hat{w}$ is orthogonal to w.

4c) It holds $\|\hat{w}\|_2^2 = \sum_{j=1}^{k} \langle u, w_j \rangle^2$.

4d) It holds $\|u - \hat{w}\|_2^2 = \|u\|_2^2 - \sum_{j=1}^{k} \langle u, w_j \rangle^2$.

For notation convenience, we will denote δ_{jr} to be the delta function such that $\delta_{ij} = 0$ if $i \neq j$, and $\delta_{ij} = 1$ if $i = j$.

Solution 4a

4a) We are searching for a minimizer of $\min_{w \in W} \|u - w\|_2$. To find such a minimizer observe that for any $w \in W$ we have $\alpha_1, \dots, \alpha_k$, such that $w = \sum_{j=1}^{k} \alpha_j w_j$ (since $w_1, \dots w_k$ is an orthonormal basis for W). Hence,

$$\|u - w\|_2^2 = \left\| u - \sum_{j=1}^{k} \alpha_j w_j \right\|_2^2 .$$

Solution 4a

4a) By the definition of norm and inner product,

$$\|u - w\|_2^2 = \Big\|u - \sum_{j=1}^{k} \alpha_k w_K\Big\|_2^2$$

$$= \|u\|_2^2 - 2\Big\langle u, \sum_{j=1}^{k} \alpha_j w_j \Big\rangle + \Big\|\sum_{j=1}^{k} \alpha_j w_j\Big\|_2^2 \qquad \text{(linearity)}$$

$$= \|u\|_2^2 - 2\sum_{j=1}^{k} \langle u, \alpha_j w_j \rangle + \sum_{i,j=1}^{k} \alpha_i \alpha_j \underbrace{\langle w_i, w_j \rangle}_{\delta_{ij}} \qquad \text{(orthogonality)}$$

$$= \|u\|_2^2 + \sum_{j=1}^{k} [\alpha_j^2 - 2\alpha_j \langle u, w_j \rangle]$$

Solution 4a

4a) We have found the equation:

$$\|u - w\|_2^2 = \|u\|_2^2 + \sum_{j=1}^{k} [\alpha_j^2 - 2\alpha_j \langle u, w_j \rangle]$$

Solution 4a

4a) We have found the equation:

$$\|u - w\|_2^2 = \|u\|_2^2 + \sum_{j=1}^{k} [\alpha_j^2 - 2\alpha_j \langle u, w_j \rangle]$$

- We see that $\|u\|_2$ is independent of the choice of w.
- Minimizing $\|u - w\|_2^2$ is equivalent to minimizing the function $f(\alpha_1, \dots, \alpha_k) = \sum_{j=1}^{k} [\alpha_j^2 - 2\alpha_j \langle u, w_j \rangle]$ over $\alpha_1, \dots, \alpha_k$.

Solution 4a

4a) continous.

- Since we can choose each α_i for $j = 1, \ldots, k$ independently, the sum is minimized by minimizing each summand.
- The function $g(\alpha) = \alpha^2 - 2\alpha c$ has its unique minimum at $\alpha = c$. Thus the sum is minimized by setting $\alpha_j = \langle u, w_j \rangle$.
- Plug in back to the equation we get that $\hat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$.

Solution 4b

4b) Let $\widehat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$ and $w = \sum_{j=1}^{k} \alpha_j w_j$. By direct computation,

Solution 4b

4b) Let $\widehat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$ and $w = \sum_{j=1}^{k} \alpha_j w_j$. By direct computation,

$$
\begin{aligned}
\langle u - \widehat{w}, w \rangle &= \langle u, w \rangle - \langle \widehat{w}, w \rangle \\
&= \sum_{j=1}^{k} \alpha_j \langle u, w_j \rangle - \sum_{i,j}^{k} \alpha_i \langle u, w_j \rangle \underbrace{\langle w_j, w_i \rangle}_{\delta_{ji}} \\
&= \sum_{j=1}^{k} \alpha_j \langle u, w_j \rangle - \sum_{j=1}^{k} \alpha_j \langle u, w_j \rangle = 0.
\end{aligned}
$$

Therefore $u - \widehat{w}$ and $w \in W$ are orthogonal. Also, $\langle u, w \rangle = \langle \widehat{w}, w \rangle$

Solution 4c

4c) Let $\widehat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$. Then

$$\begin{aligned}
\|\widehat{w}\|_2^2 &= \Big\| \sum_{j=1}^{k} \langle u, w_j \rangle w_j \Big\|_2^2 \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} \langle u, w_i \rangle \langle u, w_j \rangle \underbrace{\langle w_j, w_r \rangle}_{\delta_{ij}} \\
&= \sum_{j=1}^{k} \langle u, w_j \rangle^2.
\end{aligned}$$

Solution 4d

4d) Let $\widehat{w} = \sum_{j=1}^{k} \langle u, w_j \rangle w_j$. Then

$$\begin{aligned} \|u - \widehat{w}\|_2^2 &= \|u\|_2^2 - 2\langle u, \widehat{w} \rangle + \|\widehat{w}\|_2^2 \\ &= \|u\|_2^2 - 2 \sum_{j=1}^{k} \langle u, w_j \rangle^2 + \|\widehat{w}\|_2^2 \\ &= \|u\|_2^2 - \sum_{j=1}^{k} \langle u, w_j \rangle^2 . \end{aligned}$$