High-dimensional data analysis

Lecture 1

Review of core linear algebra concepts

Innopolis University

Fall 2018

What is this course about

- Mathematical foundations of data analysis:
    - Classical dimensionality reduction.
    - Clustering.
    - Classification.
    - Compressed sensing and embedding.

Course logistics

- 13 lectures, 13 labs
- Office hours?
- Grading:
  - Mid-term (10%) and final (20%) examinations
  - Six home assignments (60%)
  - In-class quizzes (10%)

Vectors and matrices

- Vector – element of $\mathbb{R}^n$ (or $\mathbb{C}^n$) $\begin{bmatrix} a_1 \\ \cdots \\ a_n \end{bmatrix}$

- Matrix – element of $\mathbb{R}^{n\times m}$ (or $\mathbb{C}^{n\times m}$) $\begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}$

- Unless explicitly noted, we are working with real matrices/vectors of any appropriate size.

Vector norm

- For vector space $\mathbb{R}^n$, a <u>norm</u> is
  - a function $f: \mathbb{R}^n \to [0; +\infty)$,
  - that $\forall \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n, \forall \gamma \in \mathbb{R}$:
    - $f(\boldsymbol{a} + \boldsymbol{b}) \leq f(\boldsymbol{a}) + f(\boldsymbol{b})$,
    - $f(\gamma \boldsymbol{a}) = |\gamma|\, f(\boldsymbol{a})$,
    - $f(\boldsymbol{a}) = 0 \Leftrightarrow \boldsymbol{a} = 0$.

- Common notation: $\|\boldsymbol{a}\| = f(\boldsymbol{a})$

- The norm can also be defined for $\mathbb{C}^n$

Vector norm

- Dot-product: $(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = \sum_{i=1}^{n} a_i b_i$, where $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$

- Euclidean norm (l_2): $\|\boldsymbol{a}\|_2 = \sqrt{(\boldsymbol{a}, \boldsymbol{a})} = \sqrt{\sum_{i=1}^{n} a_i^2}$
- Manhattan norm (l_1): $\|\boldsymbol{a}\|_1 = \sum_{i=1}^{n} |a_i|$
- p-norm (l_p): $\|\boldsymbol{a}\|_p = \left(\sum_{i=1}^{n} |a_i|^p\right)^{1/p}$
- ∞-norm (max): $\|\boldsymbol{a}\|_\infty = \max|a_i|$
- More norms can be invented

Metric (distance)

- For vector space $\mathbb{R}^n$, a <u>metric</u> is
  - a function $\Delta: \mathbb{R}^n \times \mathbb{R}^n \to [0; +\infty)$,
  - that $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n$:
    - $\Delta(\boldsymbol{a}, \boldsymbol{c}) \leq \Delta(\boldsymbol{a}, \boldsymbol{b}) + \Delta(\boldsymbol{b}, \boldsymbol{c})$,
    - $\Delta(\boldsymbol{a}, \boldsymbol{b}) = \Delta(\boldsymbol{b}, \boldsymbol{a})$,
    - $\Delta(\boldsymbol{a}, \boldsymbol{b}) = 0 \Leftrightarrow \boldsymbol{a} = \boldsymbol{b}$.

- We can induce metric Δ by the norm $\|\cdot\|$:

$$\Delta(\boldsymbol{a}, \boldsymbol{b}) = \|\boldsymbol{a} - \boldsymbol{b}\|$$

Metric (distance)

- If a metric in $\mathbb{R}^n$ is
  - a function $\Delta: \mathbb{R}^n \times \mathbb{R}^n \to [0; +\infty)$,
  - that $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^n, \forall \gamma \in \mathbb{R}$:
    - $\Delta(\boldsymbol{a}, \boldsymbol{c}) \leq \Delta(\boldsymbol{a}, \boldsymbol{b}) + \Delta(\boldsymbol{b}, \boldsymbol{c})$,
    - $\Delta(\boldsymbol{a}, \boldsymbol{b}) = \Delta(\boldsymbol{b}, \boldsymbol{a})$,
    - $\Delta(\boldsymbol{a}, \boldsymbol{b}) = 0 \Leftrightarrow \boldsymbol{a} = \boldsymbol{b}$,
    - $\underline{\Delta(\boldsymbol{a}, \boldsymbol{b}) = \Delta(\boldsymbol{a} + \boldsymbol{c}, \boldsymbol{b} + \boldsymbol{c}),}$
    - $\underline{\Delta(\gamma\boldsymbol{a}, \gamma\boldsymbol{b}) = |\gamma|\Delta(\boldsymbol{a}, \boldsymbol{b}).}$
- Then we can induce norm $||\cdot||$ by the metric Δ :

$$||\boldsymbol{a}|| = \Delta(\boldsymbol{a}, \boldsymbol{0})$$

Metric (distance)

- Euclidean distance (l_2)
- Manhattan distance (l_1)
- Any other metric induced by norm
- Discrete metric: $\Delta(\boldsymbol{a}, \boldsymbol{b}) = 1 \text{ if } \boldsymbol{a} = \boldsymbol{b} \text{ else } 0$
- Levenstein distance (similarity of strings)

Linear combination

- $$\sum_{i=1}^{k} a_i \boldsymbol{v}_i \text{, where } a_i \in \mathbb{R}, \boldsymbol{v}_i \in \mathbb{R}^n$$

Set of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ is called <u>linearly independent</u> if:

$$\sum_{i=1}^{k} a_i \boldsymbol{v}_i = 0 \Rightarrow \forall i \; a_i = 0$$

Eigenvectors and eigenvalues

- $$A\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad (A \in \mathbb{R}^{n\times n}, \boldsymbol{v}_i \in \mathbb{R}^n, \lambda_i \in \mathbb{C}, \boldsymbol{i} = 1 \dots n)$$
- $\det(A - \lambda I) = 0$ – characteristic polynomial
- Eigenvalues may or may not be distinct
- Distinct eigenvalues have distinct eigenvectors
- $\exists A^{-1} \Leftrightarrow \forall i: \lambda_i \neq 0$
- $\lambda_i(A^k) = \left(\lambda_i(A)\right)^k, k \in \mathbb{Z}$
- $\operatorname{tr} A = \sum \lambda_i, \ \det A = \prod \lambda_i$

Matrix norm

- Norm: a function $f: \mathbb{R}^{n\times m} \to [0; +\infty)$,
  - that $\forall A, B \in \mathbb{R}^{n\times m}, \forall \gamma \in \mathbb{R}$:
    - $f(A+B) \le f(A) + f(B)$,
    - $f(\gamma A) = |\gamma|\, f(A)$,
    - $f(A) = 0 \Leftrightarrow A = 0$.

- Usually (not necessary!) matrix norm is defined via vector norm:

$$\|A\| = \sup_{\boldsymbol{u}\neq\boldsymbol{0}} \frac{\|A\boldsymbol{u}\|}{\|\boldsymbol{u}\|}$$

Matrix norm

- $\|A\|_1 = \max\limits_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}|$
- $\|A\|_\infty = \max\limits_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}|$
- $\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2} = \sqrt{\operatorname{tr} A^{\mathrm{T}} A}$
- $\|A\|_2 = \sqrt{\lambda_{\max}(A^* A)}$

- $\|A\|_2 \le \|A\|_F$, $\|A\|_2 \le \sqrt{\|A\|_1 \|A\|_\infty}$

Matrix rank

- $\operatorname{rank} A =$ number of linearly independent columns = number of linearly independent rows.

- $A \in \mathbb{R}^{m \times n}$: $\operatorname{rank} A \leq \min(m, n)$
- $A \in \mathbb{R}^{n \times n}$: $\exists A^{-1} \Leftrightarrow \operatorname{rank} A = n$
- $\operatorname{rank} A = \operatorname{rank} \bar{A} = \operatorname{rank} A^{\mathrm{T}} = \operatorname{rank} A^{*} = \operatorname{rank} A^{*} A = \operatorname{rank} A A^{*}$

Systems of linear equations

- $$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ \qquad\qquad\quad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_n \end{cases} \Leftrightarrow A\boldsymbol{x} = \boldsymbol{b}$$

- $\operatorname{rank} A < n$: underdetermined system, infinitely many solutions
- $\operatorname{rank} A = n$: unique solution: $\boldsymbol{x} = A^{-1}\boldsymbol{b}$
- $\operatorname{rank} A > n$: overdetermined system, depends on $\boldsymbol{b}$

Matrix decompositions

- LU: $A = LU$. $A, L, U \in \mathbb{R}^{n\times n}$. L – lower triangular, U – upper triangular.
- Cholesky: $A = U^*U$. $A \in \mathbb{R}^{n\times n}$, Hermitian, positive-definite.
- QR: $A = QR$. $A \in \mathbb{R}^{m\times n}$; $Q \in \mathbb{R}^{m\times m}$, unitary; $R \in \mathbb{R}^{m\times n}$, upper triangular.
- Spectral: $A = VDV^{-1}$. $A, D, V \in \mathbb{R}^{n\times n}$, $D_{ii} = \lambda_i$, $V_{ij} = \left(\boldsymbol{v}_j\right)_i$, $\operatorname{rank} V = n$.
- SVD: $A = UDV^{\mathrm{T}}$. $A \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{m\times n}$, $D_{ii} \geq 0, D_{ij} = 0\ (i \neq j)$;
  $U \in \mathbb{R}^{m\times m}$, orthogonal; $V \in \mathbb{R}^{n\times n}$, orthogonal.