Matice ortogonální projekce a metoda nejmenších čtverců

Odpřednesenou látku naleznete v kapitole 12.4 a Dodatku C skript *Abstraktní a konkrétní lineární algebra*.

Minulá přednáška

1. Ortogonalisační proces (Gram-Schmidt).
2. Ortogonální projekce a ortogonální rejekce.
3. Ortogonální projekce na podprostor s ortogonální bází.

Dnešní přednáška

V této přednášce se zaměříme pouze na lineární prostory $\mathbb{R}^n$ nad $\mathbb{R}$.

1. Výpočet matice ortogonální projekce na podprostor $\mathrm{im}(\mathbf{A})$ dimense k v $\mathbb{R}^n$ s metrickým tensorem $\mathbf{G}$, kde $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_k)$.
2. Charakterisace matic ortogonálních projekcí v $\mathbb{R}^n$.
3. Aplikace projekcí na řešení soustav lineárních rovnic (metoda nejmenších čtverců).

Motivační úvaha

Ať $\mathbf{A} : \mathbb{R}^k \to \mathbb{R}^n$ má hodnost k a ať $\mathbf{G}$ je metrický tensor v $\mathbb{R}^n$.

1. Pokud soustava $(\mathbf{A} \mid \mathbf{b})$ řešení má, je toto řešení jediné.
   Důvod: $\mathbf{A}$ je monomorfismus (protože $\mathrm{def}(\mathbf{A}) = 0$).
2. $(\mathbf{A} \mid \mathbf{b})$ má řešení iff $\mathbf{b} \in \mathrm{im}(\mathbf{A})$.
3. Co dělat, pokud $\mathbf{b} \notin \mathrm{im}(\mathbf{A})$? Vyřešíme jinou soustavu $(\mathbf{A} \mid \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}))$.

   Proč zrovna soustavu $(\mathbf{A} \mid \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}))$?

   Protože $\mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b})$ je vektor z $\mathrm{im}(A)$, který je k vektoru $\mathbf{b}$ nejblíže ze všech vektorů z $\mathrm{im}(\mathbf{A})$.

   $(\mathbf{A} \mid \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}))$ má jediné řešení. Tomuto jedinému řešení $\widehat{\mathbf{x}}$ říkáme řešení soustavy $(\mathbf{A} \mid \mathbf{b})$ metodou nejmenších čtverců.

Závěr: potřebujeme umět spočítat $\mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b})$.

Matice ortogonální projekce na $\mathrm{im}(\mathbf{A})$, kde $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_k)$ má hodnost k

Ať $W = \mathrm{im}(\mathbf{A})$, kde $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_k) : \mathbb{R}^k \to \mathbb{R}^n$ má hodnost k. Ať $\mathbf{G}$ je metrický tensor v $\mathbb{R}^n$. Potom je matice $\mathbf{A}^T \cdot \mathbf{G} \cdot \mathbf{A}$ positivně definitní a platí[a]

$$\mathrm{proj}_W(\mathbf{x}) = \underbrace{\mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{G} \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{G}}_{\text{tzv. matice ortogonální projekce } \mathbf{P}_W} \cdot \mathbf{x}$$

[a] Tento divoký vzorec má krotkou podobu pro standardní skalární součin: platí $\mathrm{proj}_W(\mathbf{x}) = \underbrace{\mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T}_{=\mathbf{P}_W} \cdot \mathbf{x}$, protože $\mathbf{G} = \mathbf{E}_n$.

Důkaz.

Přednáška. ∎

Příklad (výpočet matice ortogonální projekce)

V prostoru $\mathbb{R}^3$ se standardním[a] skalárním součinem nalezněte matici projekce na rovinu $W = \mathrm{span}(\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix})$. Víme: pro $\mathbf{A} = \begin{pmatrix}1 & 1\\1 & 1\\1 & 0\end{pmatrix}$, je $\mathbf{P}_W = \mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T$ matice ortogonální projekce na W.

$$\mathbf{P}_W = \begin{pmatrix}1 & 1\\1 & 1\\1 & 0\end{pmatrix} \cdot \begin{pmatrix}3 & 2\\2 & 2\end{pmatrix}^{-1} \cdot \begin{pmatrix}1 & 1 & 1\\1 & 1 & 0\end{pmatrix} = \frac{1}{2} \cdot \begin{pmatrix}1 & 1 & 0\\1 & 1 & 0\\0 & 0 & 2\end{pmatrix}.$$

Například projekci vektoru $\begin{pmatrix}-20\\4\\6\end{pmatrix}$ na W spočítáme součinem

$$\frac{1}{2} \cdot \begin{pmatrix}1 & 1 & 0\\1 & 1 & 0\\0 & 0 & 2\end{pmatrix} \cdot \begin{pmatrix}-20\\4\\6\end{pmatrix} = \begin{pmatrix}-8\\-8\\6\end{pmatrix} = 6 \cdot \begin{pmatrix}1\\1\\1\end{pmatrix} + (-14) \cdot \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

[a] Metrický tensor tedy je $\mathbf{G} = \mathbf{E}_3$.

Příklad (výpočet matice ortogonální projekce)

V prostoru $\mathbb{R}^2$ se skalárním součinem s metrickým tensorem[a] $\mathbf{G} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ spočtěte matici projekce na přímku $W = \mathrm{span}(\begin{pmatrix} 1 \\ 1 \end{pmatrix})$.

Víme: pro $\mathbf{A} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, je $\mathbf{P}_W = \mathbf{A} \cdot (\mathbf{A}^T \cdot \mathbf{G} \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{G}$ matice ortogonální projekce.

Tudíž je

$$\mathbf{P}_W = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot ((1 \quad 1) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix})^{-1} (1 \quad 1) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \frac{1}{5} \cdot \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix}.$$

Například projekci vektoru $\begin{pmatrix} -6 \\ 6 \end{pmatrix}$ na W spočítáme součinem

$$\frac{1}{5} \cdot \begin{pmatrix} 2 & 3 \\ 2 & 3 \end{pmatrix} \cdot \begin{pmatrix} -6 \\ 6 \end{pmatrix} = \frac{1}{5} \cdot \begin{pmatrix} 6 \\ 6 \end{pmatrix} = \frac{6}{5} \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

[a] Jde o nestandardní skalární součin: to znamená, že jde o ortogonální projekci vzhledem ke skalárnímu součinu $\langle \mathbf{x} \mid \mathbf{y} \rangle = \mathbf{x}^T \cdot \mathbf{G} \cdot \mathbf{y}$.

Věta (charakterisace matic ortogonálních projekcí)

Ať $\mathbb{R}^n$ je vybaven skalárním součinem $\langle - \mid - \rangle$ s metrickým tensorem $\mathbf{G}$. Pro matici $\mathbf{P} : \mathbb{R}^n \to \mathbb{R}^n$ je ekvivalentní:

1. $\mathbf{P}$ je matice ortogonální projekce na lineární podprostor W dimense k.
2. $\mathrm{rank}(\mathbf{P}) = k$ a platí $\mathbf{P}^2 = \mathbf{P}$ a $\langle \mathbf{P} \cdot \mathbf{x} \mid \mathbf{y} \rangle = \langle \mathbf{x} \mid \mathbf{P} \cdot \mathbf{y} \rangle$.

Důkaz.

Přednáška. ■

Poznámka

Pro standardní skalární součin v $\mathbb{R}^n$ (tj. pro $\mathbf{G} = \mathbf{E}_n$) lze druhou podmínku přeformulovat takto:

2. $\mathrm{rank}(\mathbf{P}) = k$ a platí $\mathbf{P}^2 = \mathbf{P}$ a $\mathbf{P}^T = \mathbf{P}$.

Nalezení jediného řešení $\widehat{\mathbf{x}}$ soustavy $(\mathbf{A} \mid \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}))$

$$
\begin{aligned}
\mathbf{A}\widehat{\mathbf{x}} &= \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}) \\
&\Updownarrow \qquad \text{(vzorec pro výpočet projekce)} \\
\mathbf{A}\widehat{\mathbf{x}} &= \mathbf{A}(\mathbf{A}^T\mathbf{G}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{G}\mathbf{b} \\
&\Updownarrow \qquad \text{($\mathbf{A}$ je monomorfismus)} \\
\widehat{\mathbf{x}} &= (\mathbf{A}^T\mathbf{G}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{G}\mathbf{b} \\
&\Updownarrow \qquad \text{($\mathbf{A}^T\mathbf{G}\mathbf{A}$ je regulární)} \\
\mathbf{A}^T\mathbf{G}\mathbf{A}\widehat{\mathbf{x}} &= \mathbf{A}^T\mathbf{G}\mathbf{b}
\end{aligned}
$$

Závěr: $\mathbf{A}\widehat{\mathbf{x}} = \mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b})$ iff $\widehat{\mathbf{x}}$ je jediné řešení soustavy

$$(\mathbf{A}^T\mathbf{G}\mathbf{A} \mid \mathbf{A}^T\mathbf{G}\mathbf{b})$$

a tuto poslední soustavu lze získat jako součin matic

$$\mathbf{A}^T\mathbf{G} \cdot (\mathbf{A} \mid \mathbf{b})$$

Příklad

Tři body $\begin{pmatrix} 3 \\ 6 \end{pmatrix}$, $\begin{pmatrix} 4 \\ 9 \end{pmatrix}$, $\begin{pmatrix} 5 \\ 10 \end{pmatrix}$ v $\mathbb{R}^2$ neleží na žádné přímce tvaru $y = ax + b$.[a]

Proč? Protože soustava $(\mathbf{A} \mid \mathbf{b}) = \left(\begin{array}{cc|c} 3 & 1 & 6 \\ 4 & 1 & 9 \\ 5 & 1 & 10 \end{array}\right)$ nemá řešení (Frobeniova věta).

Vyřešme soustavu $(\mathbf{A} \mid \mathbf{b})$ metodou nejmenších čtverců (pro $\mathbf{G} = \mathbf{E}_3$).[b]

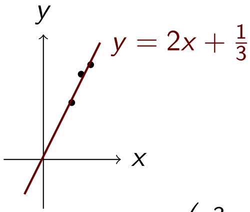
Soustava $\mathbf{A}^T \cdot (\mathbf{A} \mid \mathbf{b}) = (\mathbf{A}^T \cdot \mathbf{A} \mid \mathbf{A}^T \cdot \mathbf{b})$ má tvar $\left(\begin{array}{cc|c} 50 & 12 & 104 \\ 12 & 3 & 25 \end{array}\right)$ a má jediné řešení $\begin{pmatrix} 2 \\ 1/3 \end{pmatrix}$.

[a] Body na první pohled „téměř“ leží na přímce $y = 2x$.

[b] Řešením získáme „nejlepší možnou“ přímku, kterou lze proložit zadanými body. Říká se jí regresní přímka.

Příklad (pokrač.)

To znamená: hledaná regresní přímka má tvar $y = 2x + \frac{1}{3}$.



Vektor $\begin{pmatrix} 2 \\ 1/3 \end{pmatrix}$ není řešením soustavy $\left(\begin{array}{cc|c} 3 & 1 & 6 \\ 4 & 1 & 9 \\ 5 & 1 & 10 \end{array}\right)$.

Platí $\begin{pmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 1/3 \end{pmatrix} = \begin{pmatrix} 19/3 \\ 25/3 \\ 31/3 \end{pmatrix}$.

Čtverec chyby, které jsme se dopustili, je

$$\| \begin{pmatrix} 6 \\ 9 \\ 10 \end{pmatrix} - \begin{pmatrix} 19/3 \\ 25/3 \\ 31/3 \end{pmatrix} \|^2 = \| \begin{pmatrix} -1/3 \\ 2/3 \\ -1/3 \end{pmatrix} \|^2 = 2/3$$

Příklad (řešení soustavy rovnic metodou nejmenších čtverců, $\mathbf{G} = \mathbf{E}_3$)

Soustava $(\mathbf{A} \mid \mathbf{b}) = \left(\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 1 & 0 & 2 \end{array}\right)$ nemá řešení (Frobeniova věta).

V našem případě $\mathbf{A}^T \cdot (\mathbf{A} \mid \mathbf{b}) = (\mathbf{A}^T\mathbf{A} \mid \mathbf{A}^T\mathbf{b})$ je soustava $\left(\begin{array}{cc|c} 3 & 2 & 7 \\ 2 & 2 & 5 \end{array}\right)$ s jediným řešením $\widehat{\mathbf{x}} = \begin{pmatrix} 2 \\ 0.5 \end{pmatrix}$.

Vektor $\widehat{\mathbf{x}}$ není řešením soustavy $(\mathbf{A} \mid \mathbf{b})$, protože $\mathbf{A} \cdot \widehat{\mathbf{x}} = \begin{pmatrix} 2.5 \\ 2.5 \\ 2 \end{pmatrix}$.

Ovšem jakýkoli jiný vektor $\mathbf{x}$ by „dopadl ještě hůře“. Pro všechny vektory $\mathbf{x}$ z $\mathbb{R}^2$ totiž platí nerovnost

$$0.5 = \|\mathbf{b} - \mathbf{A} \cdot \widehat{\mathbf{x}}\|^2 \leq \|\mathbf{b} - \mathbf{A} \cdot \mathbf{x}\|^2$$

Poznámka

Pokud $(\mathbf{A} \mid \mathbf{b})$ má řešení, musí být $\mathrm{proj}_{\mathrm{im}(\mathbf{A})}(\mathbf{b}) = \mathbf{b}$. Proto lze metodu nejmenších čtverců použít i pro nalezení řešení řešitelné soustavy $(\mathbf{A} \mid \mathbf{b})$, kde $\mathbf{A}$ má lineárně nezávislé sloupce.

Příklad

Soustava $(\mathbf{A} \mid \mathbf{b}) = \left(\begin{array}{rr|r} 2 & 5 & -4 \\ -2 & 1 & -8 \\ 4 & 2 & 8 \end{array}\right)$ má (s použitím GEM) jediné řešení $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$.

Zvolte $\mathbf{G} = \mathbf{E}_3$. Pak platí $\mathbf{A}^T \cdot (\mathbf{A} \mid \mathbf{b}) = \left(\begin{array}{rr|r} 24 & 16 & 40 \\ 16 & 30 & -12 \end{array}\right)$. Tato soustava má jediné řešení $\begin{pmatrix} 3 \\ -2 \end{pmatrix}$.

Příklad (proložení paraboly, $\mathbf{G} = \mathbf{E}_n$)

Ať $\{x_1, x_2, \ldots, x_n\}$ je alespoň tříprvková množina reálných čísel. Body

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

v $\mathbb{R}^2$ leží na parabole tvaru $y = ax^2 + bx + c$ právě tehdy, když soustava rovnic

$$\left(\begin{array}{ccc|c} x_1^2 & x_1 & 1 & y_1 \\ x_2^2 & x_2 & 1 & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 & y_n \end{array}\right)$$

má řešení $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

Příklad (proložení paraboly, $\mathbf{G} = \mathbf{E}_n$, pokrač.)

Důležité pozorování: protože $\{x_1, x_2, \ldots, x_n\}$ je alespoň tříprvková množina reálných čísel, má matice soustavy

$$\left(\begin{array}{ccc|c} x_1^2 & x_1 & 1 & y_1 \\ x_2^2 & x_2 & 1 & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 & y_n \end{array}\right)$$

hodnost 3.

Opravdu: ať x_i, x_j, x_k jsou tři navzájem různé hodnoty z množiny $\{x_1, x_2, \ldots, x_n\}$.

Potom

$$\begin{vmatrix} x_i^2 & x_i & 1 \\ x_j^2 & x_j & 1 \\ x_k^2 & x_k & 1 \end{vmatrix} = (x_i - x_k) \cdot (x_i - x_k) \cdot (x_j - x_k) \neq 0$$

Příklad (proložení paraboly, $\mathbf{G} = \mathbf{E}_n$, pokrač.)

Ať $\{x_1, x_2, \ldots, x_n\}$ je alespoň tříprvková množina reálných čísel. Body

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \ldots, \begin{pmatrix} x_n \\ y_n \end{pmatrix}$$

v $\mathbb{R}^2$ lze proložit parabolu $y = ax^2 + bx + c$, kde $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ je řešením soustavy rovnic

$$\left(\begin{array}{ccc|c} x_1^2 & x_1 & 1 & y_1 \\ x_2^2 & x_2 & 1 & y_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 & y_n \end{array}\right)$$

metodou nejmenších čtverců.

Možná další (nepovinná) literatura

Řešení soustav metodou nejmenších čtverců má řadu aplikací. Je základem regresních metod v matematické statistice, viz například knihu

- Douglas C. Montgomery a George C. Runger, *Applied statistics and probability for engineers*, 3rd ed, John Wiley & Sons, New York, 2003.

pro využití metody v případě $\mathbf{G} = \mathbf{E}_n$, a například knihu

- Tilo Strutz, *Data fitting and uncertainty: A practical introduction to weighted least squares and beyond*, 2nd ed, Springer, 2016.

pro využití metody v případě obecného metrického tensoru $\mathbf{G}$.

Historická poznámka

Autorem metody nejmenších čtverců je německý matematik Karl Friedrich Gauss (1777–1855). V roce 1801 Gauss tuto metodu použil pro predikci dráhy planetky Ceres, která 40 dní po objevení[a] zmizela evropským astronomům za Sluncem. Gauss předpověděl polohu, kde se planetka za 10 měsíců opět objeví.

Viz například článek

> Donald Teets a Karen Whitehead, The discovery of Ceres: How Gauss became famous, *Mathematics Magazine* 72.2 (1999), 83–83.

nebo anglický překlad původní Gaussovy práce *Theoria Motus* z roku 1806

> Karl Friedrich Gauss, *Theory of motion of the heavenly bodies moving about the Sun in conic sections*, Dover Publications, 2004.

[a] Pravděpodobně první pozorování provedl 1. ledna 1801 italský astronom Guiseppe Piazzi.