BI-LIN, Lineární algebra – cvičení

Daniel Dombek, Luděk Kleprlík, Karel Klouda

Katedra aplikované matematiky
Fakulta informačních technologií
České vysoké učení technické v Praze

LS 2019/2020

vytvořeno: 15. února 2022, 13:46

5 Lineární kódy

5.1 Hammingova vzdálenost a řetězce nad konečnými tělesy

Kódování, v tom nejobecnějším smyslu, je přepisování řetězců znaků z jedné abecedy na jiné řetězce znaků z jiné nebo stejné abecedy. Abecedou myslíme libovolnou konečnou množinu a znaky jsou pak prvky této množiny. My budeme v tomto cvičení uvažovat výhradně abecedy totožné s konečnými tělesy $\mathbb{Z}_p$.

Konečné řetězce znaků se často nazývají slova. Množinu všech konečných neprázdných řetězců nad abecedou $\mathcal{A}$ značíme $\mathcal{A}^+$. Jako délku slova bereme počet znaků slova a množinu všech slov nad abecedou $\mathcal{A}$, která mají délku $n \in \mathbb{N}$, značíme $\mathcal{A}^n$.

Vezmeme-li za abecedu např. těleso $\mathbb{Z}_2$, značí množina $(\mathbb{Z}_2)^3$ množinu všech binárních slov délky 3:

$$000, 001, 010, 011, 100, 101, 110, 111.$$

My budeme využívat jednoznačné korespondence mezi slovem nad $\mathbb{Z}_p$ délky n a uspořádanou $n$ticí z vektorového prostoru $\mathbb{Z}_p^n$. Např. slovo 010 budeme někdy považovat za vektor $(0, 1, 0)$ ze $\mathbb{Z}_2^3$ a ten zase někdy za slovo délky 3 nad abecedou $\mathbb{Z}_2$.

Na množině stejně dlouhých slov si zavedeme **Hammingovu vzdálenost**, která je jednou z nejpoužívanějších (resp. nejzákladnějších) měr podobnosti slov:

> Pro dvě slova $u = u_1u_2 \dots u_n$ a $v = v_1v_2 \dots v_n$ stejné délky n a nad stejnou abecedou definujeme Hammingovu vzdálenost jako
>
> $$d(u, v) = \text{ počet indexů } i \in \hat{n} \text{ takových, že } u_i \neq v_i.$$

Příklad 5.1. V množině slov M najděte dva různé prvky, jejichž Hammingova vzdálenost je nejmenší.

a) $M = \{00100, 11101, 10100, 11111, 00000\}$.

b) Množina všech slov ze $\mathbb{Z}_2^5$, obsahující sudý počet jedniček.

c) Množina všech slov ze $\mathbb{Z}_2^{10}$, které vznikly spojením (zřetězením) jednoho slova ze $\mathbb{Z}_2^5$ se sebou samým.

Řešení. a) Hammingova vzdálenost je 1. Slova s touto vzdáleností jsou 00100 a 00000, 11101 a 11111, 00100 a 10100.

b) Hammingova vzdálenost dvou různých slov je 2 nebo 4. Slova se vzdáleností 2 jsou např. 11000 a 10100.

c) Min. Hammingova vzdálenost je opět dva, např. pro řetězce 0000000000 a 1000010000. □

Lineární kódy, kterými se budeme později zejména zabývat, budou definovány jako podprostory vektorových prostorů $\mathbb{Z}_p^n$ (které chápeme také jako množiny slov délky n). Prozkoumejme tedy ještě trochu slova a podprostory z těchto VP.

Příklad 5.2. Rozhodněte, jestli jsou následující množiny M podprostory zadaných vektorových prostorů V. Pokud ano, najděte matici $\mathbb{A}$ takovou, že množina M odpovídá řešení soustavy $\mathbb{A}\mathbb{x} = \theta$.

a) M je množina prvků $\mathbb{Z}_2^5$, které obsahují sudý počet jedniček.

b) M je množina prvků $\mathbb{Z}_2^n, n \in \mathbb{N}$, které obsahují sudý počet jedniček.

c) M je množina prvků $\mathbb{Z}_2^6$ takových, že počet jedniček v nich je násobek tří.

d) M je množina prvků $\mathbb{Z}_3^6$ takových, že součet jejich znaků (ciferný součet) je násobek tří.

e) M je množina prvků $\mathbb{Z}_2^6$, které vznikly spojením jednoho slova ze $\mathbb{Z}_2^3$ se sebou samým.

f) M je množina prvků $\mathbb{Z}_2^{3n}$, které vznikly spojením jednoho slova ze $\mathbb{Z}_2^n$ se sebou samým třikrát.

Řešení. Příklad a) bychom mohli vyřešit tak, že ověříme vlastnosti podprostoru, tedy uzavřenost na sčítání a násobení číslem z tělesa. Existuje ale jednodušší způsob: stačí si uvědomit, že vektor/slovo $x_1x_2x_3x_4x_5$ patří do M, právě když v $\mathbb{Z}_2$ platí[1]

$$x_1 + x_2 + x_3 + x_4 + x_5 = 0.$$

To ale není nic jiného, než homogenní soustava rovnic v $\mathbb{Z}_2^5$ a tedy M, jakožto množina jejich řešení, je podprostor. Navíc máme rovnou odpověď na druhou otázku: hledaná matice je matice

$$\mathbb{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Výsledky ostatních příkladů: b) M je podprostor a

$$\mathbb{A} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} \in \mathbb{Z}_2^{1,n}.$$

c) M není podprostor.
d) M je podprostor a

$$\mathbb{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

e) M je podprostor a

$$\mathbb{A} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

f) M je podprostor a matice $\mathbb{A}$ je s využitím blokového zápisu rovna

$$\mathbb{A} = \begin{pmatrix} \mathbb{E} & \Theta & \mathbb{E} \\ \Theta & \mathbb{E} & \mathbb{E} \end{pmatrix},$$

kde $\mathbb{E}$ je jednotková a Θ nulová matice rozměrů $n \times n$. □

[1] Takhle pod čarou připomínáme, že všechno sčítání je modulo 2 a být sudý tedy odpovídá v $\mathbb{Z}_2$ tomu, být rovný nule.

Příklad 5.3. Najděte báze podprostorů z předchozího příkladu.

Řešení. Vzhledem k tomu, že jsme našli matice homogenních soustav, jejichž je M množina řešení, je úloha najít bázi M shodná s úlohou řešit soustavy rovnic (přesněji hledat bázi S_0) a to již máme v paži.

V případě slov nad abecedami $\mathbb{Z}_p$ si zavádíme vedle Hammingovy vzdálenosti ještě **Hammingovu váhu** slova/vektoru $u \in \mathbb{Z}_p^n$, která je rovna počtu nenulových znaků. Pro Hammingovu váhu platí, že je rovna vzdálenosti od nulového vektoru:

$$\|u\| = d(u, 0^n).$$

Příklad 5.4. Dokažte následující tvrzení: Nechť P je podprostor vektorového prostoru $\mathbb{Z}_p^n$. Pak

$$\min\{d(u,v) \mid u, v \in P, u \neq v\} = \min\{\|u\| \mid u \in P, u \neq \theta\}.$$

To jest, minimální vzdálenost mezi slovy z P je právě minimální možná váha nenulového slova z P.

5.2 Lineární kódy

Lineární (n,k)-kód K je podprostor T^n dimenze k, kde T je nějaké konečné těleso (pro nás tedy $\mathbb{Z}_p$ pro nějaké prvočíslo p). Jakožto podprostor jej můžeme zadat pomocí báze (resp. generující matice, jejíž řádky jsou tato báze), nebo pomocí kontrolní matice H_K takové, že $u \in K$, právě když $H_K \cdot u = \theta$.

Poznamenejme, že u množin M v příkladu 5.2, které byly podprostory, hrály roli kontrolních matic právě hledané matice $\mathbb{A}$.

Jedním z klíčových parametrů kódu je **minimální vzdálenost**, neboť na ní závisí, kolik chyb kód objevuje resp. opravuje. Pro lineární kód K je minimální vzdálenost rovna (viz předchozí cvičení)

$$\mu(K) = \min\{d(u,v) \mid u, v \in K, u \neq v\} = \min\{\|u\| \mid u \in K, u \neq \theta\}.$$

Je-li $\mu(K) = m$, potom kód K objevuje $m-1$ (a menší) chyby a opravuje $\lfloor \frac{m-1}{2} \rfloor$ (a menší) chyby.

Z přednášky víme, že lineární (n,k)-kód je **systematický**, právě když jeho generující matici můžeme volit ve tvaru

$$G_K = \begin{pmatrix} \mathbb{E}_k & \mathbb{B} \end{pmatrix},$$

kde $\mathbb{E}_k$ je jednotková matice typu $k \times k$.

Příklad 5.5. Dokažte, že pro systematický lineární (n,k)-kód s generující maticí

$$G_K = \begin{pmatrix} \mathbb{E}_k & \mathbb{B} \end{pmatrix},$$

platí, že matice $(-\mathbb{B}^T \ \mathbb{E}_{n-k})$ je maticí kontrolní. Určete rozměry všech uvedených matic.

Příklad 5.6. Pro kód $K \subset \mathbb{Z}_2^7$ zadaný generující maticí

$$G_K = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

najděte kontrolní matici a minimální vzdálenost. Rozhodněte, jestli se jedná o systematický kód.

Řešení. Dle definice generující matice je

$$K = \langle (1,0,1,1,0,1,1), (0,1,1,1,1,0,1), (1,1,1,0,0,0,1) \rangle.$$

Jak k podprostoru najít matici H_K takovou, že tento podprostor odpovídá množině řešení homogenní soustavy s touto maticí, víme z předchozího cvičení: podprostor je také varieta a my vlastně hledáme její neparametrické rovnice!

Můžeme ale postupovat i jinak. Pokusíme se nejdříve ověřit, jestli je zadaný kód systematický. Pomocí GEM se pokusíme matici G_K upravit na matici ve tvaru

$$\begin{pmatrix} \mathbb{E} & \mathbb{B} \end{pmatrix}.$$

Jelikož GEM nemění lineární obal řádků (viz přednášku k tématu hodnost matice), jedná se stále o generující matici téhož kódu. Úprava pomocí GEM se podaří a dostaneme tak jinou generující matici

$$G'_K = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

a kód je tedy systematický. Dle předchozího cvičení je kontrolní matice rovna

$$H'_K = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

což můžeme ověřit tak, že zkontrolujeme platnost rovnice

$$H'_K (G'_K)^T = \Theta.$$

Minimální vzdálenost můžeme spočítat dvěma způsoby: buď tak, že si projdeme všechna nenulová kódová slova a spočítáme minimální Hammingovu váhu, nebo tak, že najdeme minimální počet sloupců kontrolní matice, které tvoří LZ soubor.

První způsob znamená najít všechny netriviální lin. kombinace báze. Těch je $2^3 - 1 = 7$:

$$(1,0,1,1,0,1,1), (0,1,1,1,1,0,1), (1,1,1,0,0,0,1), (1,0,0,1,1,0,0),$$
$$(0,1,0,1,0,1,0), (0,0,1,0,1,1,1), (1,1,0,0,1,1,0).$$

Toto jsou všechna kódová slova (mimo $(0,0,0,0,0,0,0)$) a minimální vzdálenost je tedy 3.

Druhý způsob vyžaduje zkušené oko dívající se na kontrolní matici: vidíme, že žádné dva sloupce nejsou stejné, což nad tělesem $\mathbb{Z}_2$ znamená, že libovolné dva sloupce tvoří LN soubor. Také si snadno všimneme, že první sloupec je součtem 4. a 5. sloupce a tedy tyto tři sloupce tvoří LZ soubor. Minimální vzdálenost kódu je tedy 3. □

Mějme lineární (n,k)-kód $K \subset\subset \mathbb{Z}_p^n$. Dekódování je zobrazení, které libovolnému slovu ze $\mathbb{Z}_p^n$ přiřadí kódové slovo s tím, že kódovému slovu u přiřadí vždy u. Jelikož se snažíme při dekódování zjistit, jaké slovo bylo původně odesláno, snažíme se dekódování definovat tak, aby nekódovému slovu, které vzniklo z kódového kvůli nějaké chybě, bylo přiřazeno nějaké nejbližší kódové slovo. Jedním ze způsobů jak to udělat, je použít tzv. standardní dekódování (viz přednáška). To stojí na násl. myšlence.

Příklad 5.7. Je-li $K \subset\subset \mathbb{Z}_p^n$ lineární (n,k)-kód, pak každá varieta W se zaměřením K má právě p^k prvků. Každé dvě takové variety jsou buďto stejné, nebo disjunktní. To znamená, že $\mathbb{Z}_p^n$ se rozkládá na variety tvaru $u + K$; každý prvek $x \in \mathbb{Z}_p^n$ leží právě v jedné z nich. Takových variet je p^{n-k}.

Zvolme v každé varietě $u + K$ slovo π_u s nejnižší možnou Hammingovou váhou, tzv. *pivot*. Pro dané u a jeho pivot π_u je tedy $u - \pi_u \in K$. Přitom podprostor K je sám o sobě varietou se zaměřením K, jeho pivotem je nulový vektor, a pro kódová slova $u \in K$ je tedy $u - \pi_u = u$.

Příklad 5.8. Pro kód $K \subset \mathbb{Z}_2^5$ zadaný generující maticí

$$G_K = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

najděte kontrolní matici a minimální vzdálenost. Rozhodněte, jestli se jedná o systematický kód a vytvořte dekódovací tabulku definující standardní dekódování.

Řešení. Z tvaru generující matice vidíme, že K je systematický kód, a dostáváme kontrolní matici

$$H_K = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

Kódová slova jsou čtyři:

$$K = \{(0,0,0,0,0), (1,0,1,0,1), (0,1,1,1,1), (1,1,0,1,0)\}.$$

Z toho je hned vidět, že minimální vzdálenost kódu je 3.

Z dřívějšího již víme, že libovolný prvek lineární variety lze zvolit jejím vektorem posunutí. Tedy všechna slova ležící ve stejné varietě jako zadané $u \in \mathbb{Z}_p^n$ nalezneme jako součty slova u s kódovými slovy $w \in K$. Například lineární varietou obsahující slovo 01110 je $\{01110, 11011, 00001, 10100\}$, jejím pivotem je slovo 00001.

Každá lineární varieta $u+K$ tedy obsahuje 4 prvky a různých lineárních variet je nutně 8. Zapíšeme je do dekódovací tabulky tak, že každá varieta tvoří jeden řádek, první řádek je přímo kód K, v prvním sloupci jsou pivoty a platí, že prvek v j-tém sloupci je součtem pivotu dané třídy a kódového slova z prvního řádku a j-tého sloupce.

00000	10101	01111	11010
00001	10100	01110	11011
00010	10111	01101	11000
00100	10001	01011	11110
01000	11101	00111	10010
10000	00101	11111	01010
00011	10110	01100	11001
01001	11100	00110	10011

Přijaté slovo 00111 potom standardně dekódujeme jako kódové slovo 01111 z prvního řádku ve stejném sloupci. Pivot 01000 ve stejném řádku a prvním sloupci říká, že přijaté slovo se od tohoto kódového slova liší právě ve druhém znaku. □

Dekódovací tabulku lze ještě zjednodušit díky definici kontrolní matice, vlastnostem lineárních variet a maticového násobení: uvažujme lineární varietu $v + K$ a libovolné slovo $u \in v + K$. Pak nutně existuje $w \in K$ takové, že $u = v + w$ a dále

$$H_K u = H_K(v + w) = H_K v + H_K w = H_K v + \theta = H_K v.$$

Vektoru $H_K u$ budeme říkat **příznak** (**syndrom**) slova u. Slova ve stejné lineární varietě $u+K$ potom mají stejný příznak! Pro každý řádek tabulky tedy ve skutečnosti stačí znát pivot a jeho příznak; přijaté slovo u pak můžeme dekódovat následovně: spočítáme jeho příznak $H_K u$, z tabulky vyčteme jeho pivot π_u, a získáme kódové slovo $u - \pi_u \in K$.

Příklad 5.9. Najděte syndromovou dekódovací tabulku pro kód z předchozího příkladu.

Řešení. Násobením kontrolní maticí počítáme příznaky pivotů z předchozí tabulky.

pivot	příznak
00000	000
00001	001
00010	010
00100	100
01000	111
10000	101
00011	011
01001	110

Kdybychom nyní chtěli standardně dekódovat slovo 00111, spočítáme

$$H_K \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Hledaný pivot je tedy dle tabulky 01000 a slovo dekódujeme jako

$$00111 - 01000 = 01111 \in K.$$

□

Příklad 5.10. Po drátě přišla následující zpráva dejvického agenta, poškozená šumem:

```
11010 11011 10101 00111 11001
01010 11010 00001 10101 11000
10101 10110 11111 10111 10001
01101 11011 10101 00001 10000
```

S použitím kódu popsaného výše dekódujte původně odeslanou zprávu. Pro zvýšení rozkoše ji pak přečtěte jako morseovku, kde `10101` je tečka, `11010` je čárka, `01111` je konec písmene a `00000` je konec slova.

Příklad 5.11. Dekódujte tutéž zprávu s použitím téhož kódu, ovšem pomocí dekódovací tabulky

00000	10101	01111	11010
00001	10100	01110	11011
00010	10111	01101	11000
00100	10001	01011	11110
01000	11101	00111	10010
10000	00101	11111	01010
01100	11001	00011	10110
00110	10011	01001	11100

Taková tabulka se od původní dekódovací tabulky liší jen volbou pivotů — například `01100` je stejně dobrým pivotem své rozkladové třídy jako `00011`, totiž oba mají minimální váhu 2. Všimněte si, že v rozkladových třídách s vahou 1 je naopak volba pivotu jednoznačná. S použitím této tabulky se zpráva dekóduje jinak; použitý kód opravuje 1-chyby, ale nikoli 2-chyby, přitom zpráva 2-chyby obsahuje.