Zápočtová písemka z Teorie informace

22. května 2024

Instrukce k vypracování: Pište stručný postup, zdůvodnění. Při výpočty s logaritmy zjednodušte vše tak, aby zůstaly jen logaritmy od lichých přirozených čísel. V případě kódování náhodné veličiny, kódujte pouze hodnoty, které mají v příslušném rozdělení nenulovou pravděpodobnost. Používat můžete psací pomůcky a oficiální tahák, který dostanete.

1. Mějme náhodné veličiny X, Y, kde sdružené rozdělení $P_{X,Y}$ je dáno tabulkou

$X \setminus Y$	1	2	3
1	$1/8$	$1/2$	$1/8$
2	$1/8$	0	$1/8$
3	0	0	0

   (a) Určete $\mathcal{H}(X)$. [1b.]

   (b) Určete $\mathcal{H}(X \mid Y)$. [2b.]

   (c) Určete $\mathcal{I}(X : Y)$. [1b.]

   (d) Nechť Z je náhodná veličina, která je rovna 1, pokud $X = Y$. Jinak je rovna 0. Určete $\mathcal{H}(Z \mid (X, Y))$. [2b.]

Řešení:

(a) $P_X \sim \left(\frac{3}{4}, \frac{1}{4}, 0\right)$

$$\mathcal{H}(X) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 2 - \frac{3}{4}\log 3.$$

(b) $P_Y \sim \left(\frac{1}{4}, \frac{3}{4}, \frac{1}{4}\right)$, $P_{X|Y=1} \sim \left(\frac{1}{2}, \frac{1}{2}, 0\right)$, $P_{X|Y=2} \sim (1, 0, 0)$, $P_{X|Y=3} \sim \left(\frac{1}{2}, \frac{1}{2}, 0\right)$

$$\begin{aligned}\mathcal{H}(X \mid Y) &= \frac{1}{4}\mathcal{H}(X \mid Y = 1) + \frac{1}{2}\mathcal{H}(X \mid Y = 2) + \frac{1}{4}\mathcal{H}(X \mid Y = 3) \\ &= 2 \cdot \frac{1}{4}\left(-2 \cdot \frac{1}{2}\log\frac{1}{2}\right) = \frac{1}{2}.\end{aligned}$$

(c) Dosadíme výsledky z předchozích bodů

$$\mathcal{I}(X : Y) = \mathcal{H}(X) - \mathcal{H}(X \mid Y) = \frac{3}{2} - \frac{3}{4}\log 3.$$

(d) Z je zcela určena veličinami (X, Y), proto $\mathcal{H}(Z \mid (X, Y)) = 0$.

Přesněji řečeno, Z je transformací (X, Y). Důsledné odvození z teorie (které jsme zde nepožadovali) by vypadalo takto: definujme $f : \{0, 1, 2\} \times \{0, 1, 2\} \to \{0, 1\}$ předpisem $f(a, b) = 1$ pokud $a = b$, jinak $f(a, b) = 0$. Potom z definice plyne $Z = f(X, Y)$. Definujme ještě $g : \{0, 1, 2\}^2 \to \{0, 1, 2\}^2 \times \{0, 1\}$ předpisem $g(a, b) = (a, b, f(a, b))$. Potom $g(X, Y) = (X, Y, Z)$. Neboli (X, Y, Z) je transformací (X, Y), a to prostou. Tedy $\mathcal{H}(Z, X, Y)$ je stejná jako $\mathcal{H}(X, Y)$ a podmíněná entropie musí být nula.

2. Mějme náhodné veličiny X, Y z předchozí úlohy. Označme $W = (X, Y)$. V následujících úlohách uvažujte pouze kódování těch hodnot veličin, které mají **nenulovou** pravděpodobnost(!):

   (a) Spočtěte entropii W. [2b.]

   (b) Vymezte délku kódových slov Shanonnova kódu pro veličinu W. [1b.]

   (c) Najděte Shanonnův kód pro veličinu W a spočtěte jeho střední délku $\mathbb{E}(|f(W)|)$. [2b.]

   (d) Najděte Huffmanův binární kód pro W a spočtěte jeho střední délku. [2b.]

   (e) Vyrovná se střední délka nějakého z kódů entropii? Bylo zřejmé z nějakých teoretických poznatků, že bude střední délka Huffmanova, či Shannonova kódu rovna/nerovna entropii? [2b.]

   (f) Pokud bychom kódovali Shannonovým kódem veličinu Y, bude střední délka kódu větší/menší/rovna střední délce Shannonova k=odu pro W? [1b.]

Řešení:

(a) $\mathcal{H}(W) = -4 \cdot \frac{1}{8}\log\frac{1}{8} - \frac{1}{2}\log\frac{1}{2} = 2.$

(b) Hodnotám $(1, 1), (2, 1), (3, 1), (1, 2), (3, 2)$ odpovídají délky slov (horní celá část logaritmu pravděpodobnosti) 3, 1, 3, 3, 3.

(c) Vzhledem k předchozímu bodu odpovídají hodnotám $(1, 1)$, $(2, 1)$, $(3, 1)$, $(1, 2)$, $(3, 2)$ například slova 000, 1, 001, 010, 011. Je třeba zachovat prefixovost! Střední délka kódu je

$$[\![f]\!]_W = \mathbb{E}(|f(W)|) = 4 \cdot 3 \cdot \frac{1}{8} + 1 \cdot \frac{1}{2} = 2.$$

(d) Jeden z možných Huffmanových kódů pro W splývá s Shannonovým kódem z předchozího bodu. Má tedy i stejnou střední délku.

(e) Střední délky Huffmanova i Shannonova kódu se entropii rovnají. To bylo jasné z toho, že pravděpodobnosti jednotlivých jevů jsou celočíselné mocniny velikosti výstupní abecedy (tedy čísla 2).

(f) Hodnoty veličiny Y mají také podobu celočíselných mocnin čísla 2. Proto střední délka Shannonova (i Huffmanova)kódu pro Y musí být rovna entropii Y. Ta je menší než entropie W. Tudíž střední délka Shannonova kódu pro Y bude menší než střední délka Shannonova kódu pro W.

3. Pan Vomáčka si volí čtyřmístný číselný PIN velmi lehkovážně, používá jen nuly a jedničky, přičemž jedničku použije **maximálně** jednu. Každá z možností je stejně pravděpodobná. Označme jednotlivé pozice v PINu jako X_i, $i = 1..4$. Následující slovní zadání vyjádři pomocí entropie, či podmíněné entropie zavedených veličin a vyřeš:

(a) Jaká je entropie celého PINu? [1b.]

(b) Jaká je entropie třetí pozice PINu? [2b.]

(c) Jaká je entropie třetí pozice PINu za předpokladu, že je druhá pozice rovna 1? [1b.]

(d) Jaká je vzájemná informace druhé a třetí pozice v PINu? [2b.]

Řešení:

(a) Celý PIN odpovídá uspořádáné čtveřici X_1, X_2, X_3, X_4. Jelikož PIN nabývá každou z hodnot 0000, 0001, 0010, 0100, 1000 se stejnou pravděpodobností, je $\mathcal{H}\left(X_1, X_2, X_3, X_4\right) = \log 5$.

(b) Pro třetí, ale také pro každou jinou, pozici platí, že je v jednom případě z pěti rovna jedné, jinak rovna nule. Tedy $P_{X_i}(0) = \frac{4}{5}$, $P_{X_i}(1) = \frac{1}{5}$. Entropie třetí pozice (a jakékoliv jiné) je tedy

$$\mathcal{H}\left(X_3\right) = -\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5} = \log 5 - \frac{8}{5}.$$

(c) Pokud je druhá pozice rovna 1, přichází v úvahu jen 1 PIN 0100. Další pozice jsou tedy touto podmínkou zcela determinovány a podmíněná entropie $\mathcal{H}\left(X_3 \mid X_2 = 0\right)$ musí být nulová.

(d) Známe entropie jednotlivých pozic, dopočteme tedy sdruženou entropii druhé a třetí pozice. Rozborem 5 možných pinů dostaneme, že uspořádaná dvojice X_2, X_3 nabývá hodnot 00, 01 a 10. První možnost nastává ve 3 případech z pěti, další dva každý jednou. Tedy

$$P_{X_2,X_3}(00) = \frac{3}{5}, \qquad P_{X_2,X_3}(01) = P_{X_2,X_3}(01) = \frac{1}{5}.$$

$$\mathcal{H}\left(X_2, X_3\right) = -\frac{3}{5}\log\frac{3}{5} - 2\cdot\frac{1}{5}\log\frac{1}{5} = \log 5 - \frac{3}{5}\log 3$$

$$\begin{aligned}\mathcal{I}\left(X_2 : X_3\right) &= \mathcal{H}\left(X_2\right) + \mathcal{H}\left(X_3\right) - \mathcal{H}\left(X_2, X_3\right) \\ &= 2\left(\log 5 - \frac{8}{5}\right) - \left(\log 5 - \frac{3}{5}\log 3\right) \\ &= \log 5 + \frac{3}{5}\log 3 - \frac{16}{5}.\end{aligned}$$

4. (a) Zakódujte jeden z PINů pana Vomáčky LZ kódem (uveďte také jasně rozbor slova). [1b.]

   (b) Zakódujte jiný z PINů pana Vomáčky frekvenčním kódem. [1b.]

Řešení:

(a) Uveďme rozbory a kódy všech PINů: Rozbory jsou $(0, 00, 0)$, $(0, 00, 1)$, $(0, 01, 0)$, $(0, 1, 00)$, $(1, 0, 00)$, příslušné odkazy na předchozí výskyty jsou $(0, 1, 0)$, $(0, 1, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(0, 0, 2)$. Koncové bity kódujeme identitou. LZ kódy jsou:

$$\begin{aligned}\mathfrak{r}(0\,00\,0) &= 0\,0\,1011\,0\,0\,0\\ \mathfrak{r}(0\,00\,1) &= 0\,0\,1011\,0\,0\,1\\ \mathfrak{r}(0\,01\,0) &= 0\,0\,1011\,1\,0\,0\\ \mathfrak{r}(0\,1\,00) &= 0\,0\,0\,1\,1011\,0\\ \mathfrak{r}(1\,0\,00) &= 0\,1\,0\,0\,1010\,0\end{aligned}$$

(b) Uveďme frekvenční kód pro všechny PINy. (Připomeňme: pozice v množinách čtveřic se stejným Parikhovým vektorem číslujeme od nuly.)

$$\begin{aligned}\mathfrak{f}(0000) &= 11010\,0\,0\\ \mathfrak{f}(0001) &= 11011\,101\,0\\ \mathfrak{f}(0010) &= 11011\,101\,1011\\ \mathfrak{f}(0100) &= 11011\,101\,1010\\ \mathfrak{f}(1000) &= 11011\,101\,10011\end{aligned}$$