

DKDA 2025Z

Lista nr 1

Autoinformacja (informacja). Jeżeli $\Pr(A)$ jest prawdopodobieństwem wystąpienia zdarzenia A , to „autoinformacja” (albo po prostu informacja) dana jest wzorem:

$$i(A) = \log_b \frac{1}{\Pr(A)} = -\log_b \Pr(A).$$

Zwykle jako podstawę logarytmu przyjmujemy $b = 2$ (i wtedy jednostką informacji jest *bit*).

Entropia. Niech niezależne zdarzenia A_i będą zbiorami wyników pewnego eksperymentu S , przy czym $\bigcup A_i = S$. Średnia informacja (inaczej *entropia*) przyporządkowana losowemu eksperymentowi dana jest wzorem:

$$H = \sum \Pr(A_i) i(A_i) = - \sum \Pr(A_i) \log_b \Pr(A_i),$$

gdzie suma liczona jest po wszystkich zdarzeniach $A_i \in S$.

Średnia długość kodu. Niech $\mathcal{A}$ będzie alfabetem składającym się z N symboli $a_1, a_2, \dots, a_N$ o prawdopodobieństwach $\Pr(a_i)$ dla $i \in \{1, 2, \dots, N\}$ oraz długościach odpowiadających im kodów: $n(a_i)$ dla $i \in \{1, 2, \dots, N\}$. Średnią długością kodu nazywamy takie l , że:

$$l = \sum_{i=1}^N \Pr(a_i) n(a_i).$$

Nierówność Krafta-McMillana. Niech $\mathcal{C}$ będzie kodem składającym się z N słów kodowych o długościach: $l_1, l_2, \dots, l_N$. Jeżeli $\mathcal{C}$ jest jednoznacznie dekodowalny, to:

$$K(\mathcal{C}) = \sum_{i=1}^N 2^{-l_i} \leq 1.$$

Zadania

Zadanie 1 Skonstruuj kody Huffmana i wylicz średnie długości kodu dla następujących ciągów występowania symboli:

- (a) $P(a) = 0.1, P(b) = 0.1, P(c) = 0.15, P(d) = 0.25$ i $P(e) = 0.4$;
- (b) $P(a) = \frac{1}{16}, P(b) = \frac{1}{8}, P(c) = \frac{1}{4}, P(d) = \frac{9}{16}$;
- (c) $P(a_1) = 0,15, P(a_2) = 0,05, P(a_3) = 0,25, P(a_4) = 0,05, P(a_5) = 0,50$.

Dla każdego przypadku sprawdź, czy zachodzi nierówność Krafta-McMillana oraz narysuj odpowiadające mu drzewo.

Zadanie 2 Wykorzystując następujące kody:

Litera	Prawdopodobieństwo	Kod 1	Kod 2
a_1	0,2	01	10
a_2	0,4	1	00
a_3	0,2	000	11
a_4	0,1	0010	010
a_5	0,1	0011	011

zakoduj ciąg $a_2a_1a_3a_2a_1a_2$. Odkoduj uzyskane ciągi bitów przy założeniu, że pierwszy bit zostanie odebrany błędnie. Jak dużo liter błędnych znajdzie się przed pierwszą poprawną dla każdego z kodów? Jak będzie wyglądała sytuacja, gdy błąd wystąpi na 3 bicie?

Zadanie 3 Oblicz średnie długości kodów dla kodów z zadania 2.

Zadanie 4 Oblicz autoinformację dla symboli z zadania 1b.

Zadanie 5 Policz entropię, dla alfabetów z zadania 1b.

Zadanie 6 Prześledzić jaką mają postać kody Huffmana dla:

(a) k jednakowo prawdopodobnych symboli;

(b) k symboli takich, że $P(a_i) = \frac{1}{2^i}$, dla $i \in \{1, \dots, k-1\}$, i $P(a_k) = \frac{1}{2^{k-1}}$.

Zadanie 7 W algorytmie Huffmana można, w zależności od kolejności kodowania symboli z jednakowym prawdopodobieństwem, uzyskać różne (nieizomorficzne) drzewa kodów. Jakiej strategii należałoby użyć, aby odchylenie standardowe długości kodów było jak najmniejsze? (Oczekiwana długość kodów będzie oczywiście jednakowa.)

Zadanie 8 Niech S oznacza określone źródło symboli, zaś S^k źródło, z którego otrzymujemy bloki k symboli z S , z tym że poszczególne elementy bloku są generowane niezależnie i każdy element bloku jest generowany zgodnie z rozkładem prawdopodobieństwa określonym dla S . Pokazać, że $H(S^k) = k \cdot H(S)$. Pokazać, że równość ta może być fałszywa, jeśli pominiemy założenie o niezależności.

Zadanie 9 Znajdź sytuacje, w których średnia długość kodu Huffmana jest:

(a) równa entropii;

(b) przybliża się do entropii plus jeden (jest jak największa).

Zadanie 10 Załóżmy, że wszystkim literom alfabetu odpowiada to samo prawdopodobieństwo. Jakie kodowanie powinno dać minimalną średnią długość kodu w tej sytuacji?

Zadanie 11 Zapisz liczby 22 i 17 za pomocą kodów Eliasa i Fibonacciego.