

Informationstheorie

ASE Übung

12.04.2019

Dezimal:

$$1234 = 1 \times 10^3 + 2 \times 10^2 + 3 \times 10^1 + 4 \times 10^0$$

Dual (binär):

$$10110 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$$

$$\Rightarrow 22$$

ASCII

| Zeichen | Dezimal | Binär |
|---------|---------|----------|
| A | 65 | 01000001 |
| B | 66 | 01000010 |
| C | 67 | 01000011 |
| ... | ... | ... |

Shannon-Entropie

$$S = - \sum_i p(x_i) \log_2 p(x_i)$$

$$\Rightarrow 0 \leq S \leq 2$$

(für 4-Symbol-Alphabet)

Entropieberechnung

ACGT...

$$\Rightarrow A = 0.35, C = 0.25, G = 0.2, T = 0.2$$

$$\Rightarrow -(0.35 \times \log_2 0.35 + 0.25 \times \log_2 0.25 + 0.2 \times \log_2 0.2 + 0.2 \times \log_2 0.2)$$

$$= 1.95887\dots$$

DNA Binärcodierung

$$2^2 = 4$$

$$\Rightarrow A = 00, C = 01, G = 10, T = 11$$

$$\Rightarrow 1 - \left(\frac{2}{8}\right) = 0.75 \text{ Platzersparnis}$$

Menschliches Genom

$$\begin{aligned} & 2 \times 3 \times 10^9 \text{ Bits} \\ &= \frac{6}{8} \times 10^9 \text{ Bytes} \\ &= \frac{6}{8} \times 10^3 \text{ MB} = 750 \text{ MB} \end{aligned}$$

Frage 1

Mit welcher einfachen Formel kann die für eine optimale Codierung benötigte Anzahl von Bits eines Alphabets mit A Zeichen ermittelt werden?

Frage 1

$$nB = \lceil \log_2 A \rceil$$

Protein Binärcodierung

$$\Rightarrow \lceil \log_2 20 \rceil = 5$$

Frage 2

Gegeben sei die Codierung $A=00$, $C=01$, $G=10$, $T=11$. Mit welcher einfachen Bitoperation kann die jeweils komplementäre Base bestimmt werden?

Frage 2

A=00, C=01, G=10, T=11



A=11, C=10, G=01, T=00

Sequenz 1

$$-\underbrace{(0.2 \log_2 0.2)}_A + \underbrace{0.15 \log_2 0.15}_C + \underbrace{0.35 \log_2 0.35}_G + \underbrace{0.3 \log_2 0.3}_T$$
$$= 1.92612$$

Sequenz 2

$$-\underbrace{(0.075 \log_2 0.075)}_A + \underbrace{0.425 \log_2 0.425}_C + \underbrace{0.4 \log_2 0.4}_G + \underbrace{0.1 \log_2 0.1}_T$$
$$= 1.66588$$

Sequenz 3

$$-\underbrace{(0 \log_2 0)}_A + \underbrace{0.55 \log_2 0.55}_C + \underbrace{0.45 \log_2 0.45}_G + \underbrace{(0 \log_2 0)}_T$$
$$= 0.99277$$

Frage 3

Im Genom beobachten Sie Abschnitte mit sehr hohen Entropiewerten (nahe 2) zwischen längeren Abschnitten mit geringerer Entropie. Was könnte dies bedeuten?