Конечная случайная схема и энтропия. Энтропия пересечения и условная энтропия. Количество информации. Код Хэмминга

Руслан Назирович Мокаев

Математико-механический факультет,
Санкт-Петербургский государственный университет

Санкт-Петербург, 27.02.2024

Содержание лекции

- Напоминалка и необходимая информация
- Конечная случайная схема
- Энтропия и условная энтропия
- Количество информации. Свойства. Примеры
- Код Хэмминга

Напоминалка

Пусть S – конечное множество, $|S| = n$.
Пусть задана функция $f : S \to [0,1], \forall \omega \in S\ \exists!\ f(\omega) \in [0,1]$
$\sum_{\omega \in S} f(\omega) = 1$. Определим $\forall A \subseteq S$ величину $Pr(A) = \sum_{\omega \in A} f(\omega)$

Функция f в общем то и не нужна. Достаточно иметь Pr.

Определение: (S, Pr) называется вероятностным пространством.
S – пространство элементарных событий.
$\omega \in S$ – элементарное событие (исход), $A \subseteq S$ – событие.
$Pr(A)$ – вероятность A.
$A, B \subseteq S,\ Pr(A \cap B) = 0$ – несовместные события.

Свойства вероятности:

- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- $Pr(A) + Pr(S \setminus A) = 1$
- $Pr(A \cup B) \leq Pr(A) + Pr(B)$
- $Pr(A) = Pr(A \setminus B) + Pr(A \cap B)$

Неравенство Йенсена

Определение: функция f называется **выпуклой** на $X \in \mathbb{R}$, если $\forall x_1, x_2 \in X$ и $\forall \alpha \in [0,1]$ выполняется неравенство $f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$

Нер-во Йенсена: Пусть f – выпуклая на X функция. Тогда $f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i)$, где $x_i \in X$, $\alpha_i \geq 0$, $\sum_{i=1}^{n} \alpha_i = 1$.

Док-во: База при $n = 2$ верна по определению выпуклой функции. Пусть верно для n. Тогда

$$f\left(\sum_{i=1}^{n+1} \alpha_i x_i\right) = f\left((1-\alpha_{n+1}) \sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} x_i + \alpha_{n+1} x_{n+1}\right) \leq$$

$$\leq (1-\alpha_{n+1}) f\left(\sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} x_i\right) + \alpha_{n+1} f(x_{n+1}) \leq$$

$$\leq (1-\alpha_{n+1}) \sum_{i=1}^{n} \frac{\alpha_i}{1-\alpha_{n+1}} f(x_i) + \alpha_{n+1} f(x_{n+1}) = \sum_{i=1}^{n+1} \alpha_i f(x_i)$$

Конечная случайная схема и энтропия

Определение: Пусть $A_1, A_2, \ldots, A_n$ – разбиение множества исходов S вероятностного пространства (S, Pr).
Конечной случайной схемой (КСС) называется схема α, сопоставляющая каждому A_i вероятность $Pr(A_i)$.

Пример: $\begin{pmatrix} A_1 & A_2 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ и $\begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix}$. Исходы экспериментов?

Пример: В случае правильной игральной кости множеством исходов будет $A_i = i, Pr(A_i) = \frac{1}{6}$ и КСС будет $\begin{pmatrix} A_1 & A_2 & \ldots & A_6 \\ \frac{1}{6} & \frac{1}{6} & \ldots & \frac{1}{6} \end{pmatrix}$.

Определение: **Собственная информация** α – случайная величина, определяемая формулой $I(A_i) = -\log Pr(A_i)$.

Определение: **Энтропией КСС** называется величина

$$H(\alpha) := -\sum_{i=1}^{n} Pr(A_i) \cdot \log Pr(A_i)$$ (формула Шеннона)

Интуиция для понимания понятия энтропия

У нас есть курицы двух видов: оранжевые и синие. Имеется три курятника:

1. A: 6 оранжевых и 1 синяя;
2. B: 1 оранжевая и 10 синих;
3. C: 5 оранжевых и 5 синих.

Опишите ваше "удивление" при извлечении курицы из каждого курятника.

Оказывается, что ваше удивление некоторым образом обратно пропорционально вероятности извлечь курицу определенного цвета из курятника. Например, вероятность извлечь синюю курицу из курятника A мала, а 'удивление' от ее извлечение будет большим!
А как измерять "удивление"?

Есть соблазн использовать формулу Удивление $= \frac{1}{\text{Вероятность}}$.

Рассмотрим монету $Pr\{\text{О}\} = 1, Pr\{\text{Р}\} = 0$. Удивление от выпадения орла должно равняться 0! А у нас 1. Давайте будем использовать $\log$!

Рассмотрим другую монету $Pr\{\text{О}\} = 0.9, Pr\{\text{Р}\} = 0.1$. Удавление от выпадения орла $= 0.15$, от решки $= 3.32$. Makes sense!

Бросим монетку три раза. Какова вероятность получить OOP? Удивление находим по формуле $\log\left(\frac{1}{0.9 \cdot 0.9 \cdot 0.1}\right) = -\log(0.9 \cdot 0.9 \cdot 0.1) =$ $= -2 \cdot \log(0.9) - 1 \cdot \log(0.1)$.

Т.е. удивление равно суммарному удивлению каждого отдельного броска.

А после 100 бросков?

$$\text{Удивление} = \underbrace{(0.9 \cdot 100)}_{\substack{\text{Ожидаемое} \\ \text{количество орлов}}} \cdot 0.15 + \underbrace{(0.1 \cdot 100)}_{\substack{\text{Ожидаемое} \\ \text{количество решек}}} \cdot 3.32.$$

Как найти "среднее удивление" от одного броска такой монеты?

Среднее удивление $= 0.9 \cdot 0.15 + 0.1 \cdot 3.32 = 0.47$.

Таким образом, энтропия монеты – это среднее удивление, которое мы получаем от одного броска монеты! Иными словами: энтропия – это математическое ожидание величины 'удивления'.

Посчитаем энтропии для наших курятников. Энтропия будет количественной характеристикой похожести (или различия) количеств оранжевых и синих куриц!

Свойства энтропии

- $H(\alpha) \geq 0$.
- Энтропия характеризует **неопределенность** (мера неопределенности), заключенную в КСС.
- Для любой α с k исходами справедливо $H(\alpha) \leq \log k$.

Док-во: $f(x) := -x \cdot \log x$. На $[0,1]$ функция $f(x)$ строго вогнутая $\Rightarrow$ по неравенству Йенсена $\sum_{i=1}^{n} \lambda_i \cdot f(x_i) \leq f(\sum_{i=1}^{n} \lambda_i \cdot x_i)$, причем равенство достигается только когда $x_1 = \ldots = x_n$.

Тогда возьмём $x_i = Pr(A_i)$ и $\lambda_i = \frac{1}{k}$ $\forall i \in \overline{1:k}$, получаем $\sum_{i=1}^{k} \frac{1}{k}(-Pr(A_i) \cdot \log Pr(A_i)) \leq -\sum_{i=1}^{k} \frac{1}{k} Pr(A_i) \cdot \log(\sum_{i=1}^{k} \frac{1}{k} Pr(A_i))$

$-\frac{1}{k} \sum_{i=1}^{k} Pr(A_i) \cdot \log Pr(A_i) \leq -\frac{1}{k} \log \frac{1}{k}$

$-\sum_{i=1}^{k} Pr(A_i) \cdot \log Pr(A_i) \leq \log k$

max энтропию имеет КСС с k равновероятностными исх-ми.

$H(\alpha) = 0 \Leftrightarrow \exists!$ достоверный исход в α (с вероятностью 1 выбирается один и тот же символ).

Энтропия пересечения и условная энтропия

Определение: Пусть есть КСС α с исходами $A_1, \ldots, A_k$ и КСС β с исходами $B_1, \ldots, B_l$. Их пересечением $\alpha \cap \beta$ называют КСС, исходы которой это $A_i \cap B_j$ $\forall i \in \overline{1:k}, j \in \overline{1:l}$.

Тогда $H(\alpha \cap \beta) = -\sum\limits_{i=1}^{k}\sum\limits_{j=1}^{l} Pr(A_i \cap B_j) \cdot \log Pr(A_i \cap B_j)$.

Т.к. $Pr(A_i \cap B_j) = Pr(A_i) \cdot Pr(B_j|A_i) \Rightarrow H(\alpha \cap \beta) =$

$$= -\sum_{i=1}^{k}\sum_{j=1}^{l} Pr(A_i) \cdot Pr(B_j|A_i) \cdot (\log Pr(A_i) + \log Pr(B_j|A_i)) =$$

$$= -\sum_{i=1}^{k}\sum_{j=1}^{l} Pr(A_i) \cdot Pr(B_j|A_i) \cdot \log Pr(A_i) -$$

$$-\sum_{i=1}^{k}\sum_{j=1}^{l} Pr(A_i) \cdot Pr(B_j|A_i) \cdot \log Pr(B_j|A_i) =$$

$$= -\sum_{i=1}^{k} Pr(A_i) \cdot \log Pr(A_i) \cdot \sum_{j=1}^{l} Pr(B_j|A_i) +$$

$$+ \sum_{i=1}^{k} Pr(A_i) \cdot \left(-\sum_{j=1}^{l} Pr(B_j|A_i) \cdot \log Pr(B_j|A_i) \right) =$$

$$= -\sum_{i=1}^{k} Pr(A_i) \cdot \log Pr(A_i) + \ldots = H(\alpha) + \ldots$$

Определение: Величину $H(\beta|A_i) := -\sum_{j=1}^{l} Pr(B_j|A_i) \cdot \log Pr(B_j|A_i)$ называют **условной энтропией** β **при условии** A_i.

Определение: Величину $H_\alpha(\beta) := \sum_{i=1}^{k} Pr(A_i) \cdot H(\beta|A_i)$ называют (средней) **условной энтропией** β **в схеме** α.

Таким образом, $H(\alpha \cap \beta) = H(\alpha) + H_\alpha(\beta)$.

Докажем, что $0 \le H_\alpha(\beta) \le H(\beta)$.

Неотрицательность следует из неотрицательности энтропий.

fix j, $f(x) = -x \cdot \log x$, $\lambda_i = Pr(A_i), x_i = Pr(B_j|A_i)\ \forall i \in \overline{1:k}$

$$\text{Нер-во Йенсена:}\quad \sum_{i=1}^{k} Pr(A_i) \cdot \big(-Pr(B_j|A_i) \cdot \log Pr(B_j|A_i)\big) \le$$

$$\le \left(-\sum_{i=1}^{k} Pr(A_i) \cdot Pr(B_j|A_i)\right) \cdot \log \sum_{i=1}^{k} Pr(A_i) \cdot Pr(B_j|A_i)$$

$$\text{ПЧ} = \left(-\sum_{i=1}^{k} Pr(A_i) \cdot Pr(B_j|A_i)\right) \cdot \log \sum_{i=1}^{k} Pr(A_i) \cdot Pr(B_j|A_i) =$$

$$= -\left(\sum_{i=1}^{k} Pr(B_j \cap A_i)\right) \cdot \log \sum_{i=1}^{k} Pr(B_j \cap A_i) = -Pr(B_j) \cdot \log Pr(B_j)$$

Просуммируем по j:

$$\sum_{j=1}^{l}\sum_{i=1}^{k} Pr(A_i)\cdot\big(-Pr(B_j|A_i)\cdot\log Pr(B_j|A_i)\big) \le \sum_{j=1}^{l}\big(-Pr(B_j)\cdot\log Pr(B_j)\big)$$

$$\sum_{i=1}^{k} Pr(A_i)\cdot\sum_{j=1}^{l}\big(-Pr(B_j|A_i)\cdot\log Pr(B_j|A_i)\big) \le -\sum_{j=1}^{l} Pr(B_j)\cdot\log Pr(B_j)$$

$$\sum_{i=1}^{k} Pr(A_i)\cdot H(\beta|A_i) \le H(\beta) \Rightarrow H_\alpha(\beta) \le H(\beta)$$

$H_\alpha(\beta) = H(\beta) \Leftrightarrow$ все $Pr(B_j|A_i)$ равны между собой.

Ф-ла полной вероятности $\forall j \in \overline{1:l}\ Pr(B_j) = \sum_{r=1}^{k} Pr(B_j|A_r)\cdot Pr(A_r)$

$$\forall j \in \overline{1:l}\ Pr(B_j) = Pr(B_j|A_i)\cdot\sum_{r=1}^{k} Pr(A_r) = Pr(B_j|A_i)$$

То есть $\forall i \in \overline{1:k}, j \in \overline{1:l}\ Pr(B_j) = Pr(B_j|A_i)$

Определение: События A и B – **взаимно независимы** $\Leftrightarrow Pr(A \cap B) = Pr(A) \cdot Pr(B) \Leftrightarrow Pr(A) \cdot Pr(B|A) = Pr(A) \cdot Pr(B) \Leftrightarrow Pr(B|A) = Pr(B)$.

Определение: КСС α и β называются **независимыми**, когда все исходы α независимы со всеми исходами β.

Замечание: Если α и β независимы, то $H_\alpha(\beta)$ максимальна и равна $H(\beta)$.

Взаимная информация

Определение: Величина $I(\alpha,\beta) = H(\beta) - H_\alpha(\beta)$ называется **взаимной информацией** между схемами α и β.

Свойства:

- $I(\alpha,\beta) \geq 0$
- $I(\alpha,\beta) = H(\beta) \Leftrightarrow H_\alpha(\beta) = 0$ (отображения полностью зависимы, одно определяет другое и наоборот)
- $I(\alpha,\beta) = I(\beta,\alpha)$
- $I(\alpha,\beta) = 0 \Leftrightarrow \alpha$ и β независимы.

Пример:

Загадано натуральное число $x \in \overline{1:N}$
β – опыт, состоящий в нахождении x, β_m – опыт, показывающий, делится ли x на m, $m \in \overline{1:N}$.
У β есть N исходов, у β_m – два исхода.

$$H_{\beta_m}(\beta) = Pr(x \vdots m) \cdot H(\beta | x \vdots m) + Pr(m \nmid x) \cdot H(\beta | m \nmid x)$$

$q := \lfloor \frac{N}{m} \rfloor$ – количество чисел от 1 до N, делящихся на m. Тогда
$Pr(x \vdots m) = \frac{q}{N}$, $Pr(m \nmid x) = \frac{N-q}{N}$.

$$H(\beta|x \vdots m) = -\sum_{i \,\vdots\, m,\ i\in\overline{1:N}} \frac{1}{q}\cdot\log\frac{1}{q} = -\frac{q}{q}\cdot\log\frac{1}{q} = \log q$$

Аналогично $H(\beta|m \nmid x) = \log(N-q) \Rightarrow H_{\beta_m}(\beta) = \frac{q}{N}\cdot\log q + \frac{N-q}{N}\cdot\log(N-q)$

$$\begin{aligned} I(\beta_m,\beta) &= \log N - \frac{q}{N}\cdot\log q - \frac{N-q}{N}\cdot\log(N-q) = \\ &= \frac{q}{N}\cdot\log N - \frac{q}{N}\cdot\log q - \frac{N-q}{N}\cdot\log N - \frac{N-q}{N}\cdot\log(N-q) = \\ &= -\frac{q}{N}\cdot\log\frac{q}{N} - \frac{N-q}{N}\cdot\log\frac{N-q}{N} \le \log 2 \end{aligned}$$

Равенство достигается при $q = N - q = \frac{N}{2}$.

Данетки

Условие: Загадано число от 1 до N. Есть возможность задавать бинарные вопросы (получать ответ 'да' или 'нет'). Сколько бинарных вопросов необходимо задать, чтобы гарантировано отгадать число?
Опыт β – угадать число;
Опыт α – задать любой общий (да/нет) вопрос и получить ответ.

$H(\beta) = \log N$ (числа загаданы с равной вероятностью)
$H(\alpha) \leq \log 2$ (поскольку есть всего 2 варианта ответа)

$H(\alpha_1 \alpha_2 \ldots \alpha_k) \leq \log 2^k = k \log 2$ (k вопросов, 2 варианта ответа)

Чтобы угадать число потребуется $k \geq \frac{\log N}{\log 2} = \log_2 N$ вопросов

Есть ли алгоритм, который умеет угадывать загаданное число за $O(\log N)$?

Избыточное кодирование

Есть сообщение $u \in \{0,1\}^k$, которое нужно передать.
Можем передавать сообщение $x(u) \in \{0,1\}^n, n \geq k$, содержащую некоторую избыточную информацию (канал связи шумит и может допускать ошибки), но не более d ошибок на сообщение.

β заключается в нахождении всех d ошибок. Сколько у β исходов? Для каждого количества ошибок j от 0 до d есть $\binom{n}{j}$ вариантов их расположения, то есть всего исходов у β ровно $\displaystyle\sum_{j=0}^{d}\binom{n}{j}$.

Следовательно, $H(\beta) = \log \displaystyle\sum_{j=0}^{d}\binom{n}{j}$

α – дополнительное сообщение размера $n-k$. Их 2^{n-k} и
$\Rightarrow H(\alpha) = \log 2^{n-k} = (n-k)\log 2.$

Чтобы гарантировано найти все ошибки нужно $H(\alpha) \geq H(\beta)$

$$(n-k)\log 2 \geq \log \sum_{j=0}^{d} \binom{n}{j} \Rightarrow n-k \geq \log_2 \sum_{j=0}^{d} \binom{n}{j} \Rightarrow$$

$$\Rightarrow k \leq n - \log_2 \sum_{j=0}^{d} \binom{n}{j}$$

Т.о., если канал связи допускает не более d ошибок, то для передачи сообщения размера k понадобится не менее $k + \log_2 \sum_{j=0}^{d} \binom{n}{j}$ бит.

Или, поскольку количество ошибок обычно зависит от размера переданного сообщения, если передаётся n бит и из них не более d могут быть ошибочными, то в переданном сообщении можно закодировать сообщение длиной не более $n - \log_2 \sum_{j=0}^{d} \binom{n}{j}$.

Код Хэмминга

Предыдущая задача при $d = 1$. Известно, что $2^{n-k} \geq \sum_{j=0}^{1} \binom{n}{j} = 1 + n$.
$l := n - k$ – длина "избыточного" сообщения. Тогда $k \leq 2^l - l - 1$.

l	k
1	0
2	1
3	4
4	11
5	26
6	57

Чем больше сообщение, тем относительно меньше нужно лишней информации.
Как передавать дополнительную информацию?

Пример: Пусть $k = 12$ и мы хотим передать сообщение $u = 101101011100$

$$A = \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|c|}
\hline
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\hline
0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
\hline
0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\hline
\end{array}$$

Зарезервируем в сообщении длины 17 места с номерами 2^i $(1, 2, 4, 8, 16)$, а на остальные позиции запишем сообщение: $x_0(u) = \_\_1\_011\_0101110\_0$

Подберём на позицию 2^i такую цифру, чтобы произведение $x(u)$ и i-й строки матрицы было равно 0.
На «неопределенных» поз-ях в строке с номером i стоят 0 $(2^j = 10\dots0)$.
На позиции 2^i в i-й строке стоит 1.

$?*1+\_*0+1*1+\_*0+0*1+1*0+1*1+\_*0+0*1+1*0+0*1+1*0+1*1+1*0+0*1+\_*0+0*1$
$=?+1+1+1=1+?=0 \Rightarrow ?=1.$

Получается $x_1(u) = 1\_1\_011\_0101110\_0$

Аналогично делаем для остальных. Итого $x(u) = 111101100101110000$

Как определять позицию ошибки? $y = 111101100001110000$

Посчитаем $A \times y^T = (0,1,0,1,0)^T$ – двоичная запись позиции с ошибкой
Старший бит – справа. Почему так?

При умножении на i-ю строку матрицы j-я позиция сообщения влияла только если $A[i,j] = 1$, то есть если на i-м месте в двоичной записи числа j стояла $1 \Rightarrow$ результат произведения строки матрицы на столбец сообщения изменился (став 1) только для тех строк, где на 10-й позиции стояла 1 (а это строки с номерами, равными позициям, где в двоичной записи числа 10 стоят 1), а для остальных строк остался 0.