Энтропия на сообщение

4 января 2023 г. 4:44

Одномерная энтропия источника
Энтропия символа x_t (из ансамбля $X_t = X$), сгенерированного в момент времени t, не зависит от t ($H(X_t) = H(X)$) и называется одномерной энтропией источника (процесса). Обозначим её как $H_1(X)$.
$H_1(X)$ не учитывает зависимость между символами, порожденными источником

n-мерная энтропия (для блока длиной n символов)

Рассмотрим $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ из $X_1 X_2 \dots X_n = X^n$.
Энтропия $H(X_1 X_2 \dots X_n) = H(X^n)$ называется n-**мерной энтропией** процесса.
Энтропия на символ для последовательности длины n определяется как:

$$H_n(X) = \frac{H(X^n)}{n},$$

энтропия последовательности из n символов / n

Другой способ:

$$H(X_n | X_1, \dots, X_{n-1}) = H(X|X^{n-1}).$$

или вычисляем как условную энтропию

Энтропия на сообщение:

$$\lim_{n\to\infty} H_n(X) \text{ и } \lim_{n\to\infty} H(X|X^n)$$

тоже два способа

Theorem

Для дискретного стационарного процесса (источника)

A. $H(X|X^n)$ *не возрастает с увеличением n;*
B. $H_n(X)$ *не возрастает с увеличением n;*
C. $H_n(X) \geq H(X|X^{n-1})$;
D. $\lim_{n\to\infty} H_n(X) = \lim_{n\to\infty} H(X|X^n)$.

Доказательство---
A.
Следует из свойства условной энтропии о том, что если растет количество условий, энтропия либо остается такая же (при независимых событиях), либо уменьшается
C.

$$H(X^n) = H(X) + H(X|X^1) + \dots + H(X|X^{n-1}) \geq$$
$$\geq \quad nH(X|X^{n-1}) \geq nH(X|X^n)$$

расписано по определению

этот член – самый маленький в сумме, поэтому $\geq$

$H(X|X^n) < H(X|X^{n-1})$ поэтому снова $\geq$

Т.е. получаем
$H(X^n) \geq nH(X|X^n)$
Поделим левую и правую часть на n
(вспомним формулу $H_n(X) = \dfrac{H(X^n)}{n}$)
Получим:
$H_n(X) \geq H(X|X^n)$
ч.т.д.
B.
Запишем энтропию блока длиной n+1 символов $H(X^{n+1})$ в развернутом виде:
$H(X^{n+1}) = H(X_1 \dots X_n X_{n+1})$
Разобьем на сумму энтропии всего сообщения длиной n символов и энтропии последнего элемента при условии известных предыдущих:
$H(X^{n+1}) = H(X_1 \dots X_n X_{n+1}) = H(X_1 \dots X_n) + H(X_{n+1}|X_1, \dots, X_n)$
По определению $H(X_1 \dots X_n) = H(X^n) = nH_n(X)$
Из доказанного выше свойства C следует:
$nH(X_{n+1}|X_1, \dots, X_n) \leq H(X^n)$
$H(X_{n+1}|X_1, \dots, X_n) \leq H_n(X)$
Поэтому $H(X^{n+1}) \leq nH_n(X) + H_n(X)$
В правой части вынесем $H_n(X)$ за скобку, получаем $(n+1)H_n(X)$
$H(X^{n+1}) \leq (n+1) H_n(X)$
Разделим правую и левую части на n+1
$\dfrac{H(X^{n+1})}{n+1} \leq H_n(X)$
Левую часть можно еще по определению записать как $H_{n+1}(X)$
$H_{n+1}(X) \leq H_n(X)$
ч.т.д.
D.
1.
Энтропию на сообщение можно выразить двумя способами, как описано выше. Докажем, что эти две выраженные по-разному величины равны
По свойству энтропия не может быть отрицательной, поэтому $H_n(X)$ и $H(X|X^n)$ >=0
Также они не возрастают (по свойствам A и B)
Поэтому существуют пределы $\lim_{n\to\infty} H_n(X)$ и $\lim_{n\to\infty} H(X|X^n)$
Из свойства C следует:
$\lim_{n\to\infty} H_n(X) \geq \lim_{n\to\infty} H(X|X^n)$
2.
Для m<n:

Расписываем энтропию блока из n символов по определению, затем берем некое m, которое меньше n, и расписываем через m и n:

$H\left(X^n\right)=H\left(X_1 \ldots X_n\right)=H\left(X_1 \ldots X_m\right)+H\left(X_{m+1} \mid X_1, \ldots, X_m\right)+\ldots+H\left(X_n \mid X_1, \ldots, X_{n-1}\right)$

По определению $H\left(X_1 \ldots X_m\right)=m H_m(X)$. Это энтропия блока из m символов

Получается, остальные слагаемые - для символов с m по n, то есть для последних n-m символов

$H\left(X_{m+1} \mid X_1, \ldots, X_m\right)=H\left(X \mid X^m\right)$

В сумме это самая большая энтропия из тех, которые для последних n-m символов, все следующие - меньше

Получается, если умножить ее на (n-m), получим значение больше, чем сумма энтропий для последних n-m символов

Таким образом:

$H\left(X^n\right) \leq m H_m(X)+(n-m) H\left(X \mid X^m\right)$

Разделим левую и правую части на n и запишем предел ($\frac{m}{n}$ в пределе обратится в 0)

$\lim_{n \rightarrow \infty} H_n(X) \leq H\left(X \mid X^m\right)$ для любого m

Устремим m к бесконечности:

$\lim_{n \rightarrow \infty} H_n(X) \leq \lim_{m \rightarrow \infty} H\left(X \mid X^m\right)$

ИТОГ 1 и 2

Из 1 следует, что <=, а из второго - что >=, поэтому 1 и 2 существуют одновременно, если =

Таким образом, $\lim_{n \rightarrow \infty} H_n(X)=\lim_{n \rightarrow \infty} H\left(X \mid X^n\right)$

Обозначим $H_{\infty}(X)=\lim_{n \rightarrow \infty} H_n(X)$ и $H\left(X \mid X^{\infty}\right)=\lim_{n \rightarrow \infty} H\left(X \mid X^n\right)$, тогда $H_{\infty}(X)=H\left(X \mid X^{\infty}\right)$

Отсюда можно сделать два вывода. Есть два способа, которыми мы можем кодировать данные (если есть зависимость от предыдущих символов, иными словами, память):

1. Можно расширять алфавит
   Например, есть сообщение abcd, его алфавит {a, b, c, d}. Создадим алфавит со всевозможными подстроками длины 2: {aa, ab, ac, ad, ba, bb, bc, bd, ca, cb, cc, cd, da, db, dc, dd}. Таким образом можно кодировать блоками по 2 символа
2. Можно учитывать контекстную зависимость текущей буквы от предыдущих n букв

Если рассматриваем **дискретный стационарный источник без памяти** (дискретный ансамбль), то:

1. Энтропия для блока = сумма энтропий символов этого блока
2. Энтропия блока = кол-во символов * энтропия одного символа
3. Энтропия на символ = энтропия одного символа
4. Энтропия на символ при количестве символов, устремленном к бесконечности = энтропия одного символа
5. Условная энтропия символа при условии известных n предыдущих символов = энтропия одного символа (т.к. нет зависимости от предыдущих символов)
6. Условная энтропия символа при известных всех предыдущих = энтропия одного символа

- $H(X_1 ... X_n) = H(X_1) + ... + H(X_n)$.
- $H(X^n) = nH(X)$.
- $H_n(X) = H(X)$,
- $H_\infty(X) = H(X)$.
- $H(X|X^n) = H(X_{n+1}|X_1, ..., X_n) = H(X)$,
- $H(X|X^\infty) = H(X)$.

Но отсюда не следует, что для такого источника нужно кодировать каждую букву независимо от других

Если рассматриваем **марковский источник**:

- $H(X|X^n) = H(X_{n+1}|X_1, ..., X_n) = H(X_{n+1}|X_{n-s+1}, ..., X_n) = H(X|X^s)$
- $H(X|X^\infty) = H(X|X^s)$.
- $H(X^n) = H(X_1 \ldots X_s X_{s+1} \ldots X_n)$
  $$= H(X_1 \ldots X_s) + H(X_{s+1} \ldots X_n | X_1, \ldots, X_s).$$
- $H(X_{s+1} \ldots X_n | X_1, \ldots, X_s) = H(X_{s+1}|X_1, \ldots, X_s) +$
  $$\begin{aligned} &+ \quad H(X_{s+2}|X_2, \ldots, X_{s+1}) + \ldots \\ &+ \quad H(X_n|X_{n-s}, \ldots, X_{n-1}) \\ &= \quad (n-s)H(X|X^s). \end{aligned}$$
- $\frac{H(X^n)}{n} = \frac{sH_s(X)}{n} + \frac{(n-s)H(X|X^s)}{n}$