Дискретный ансамбль. Собственная информация и её свойства. Энтропия дискретного ансамбля и её свойства

3 января 2023 г. 18:39

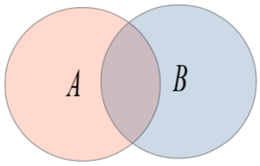
Дискретный ансамбль (X = {x}) - дискретное множество, содержащее конечное число элементов (элементарных событий) x ∈ X
Для каждого элементарного события определена вероятность p(x) >= 0
Множество чисел { p(x) } задаёт распределение вероятностей. Сумма элементов этого множества = 1
Пара "элементарное событие и её вероятность" называется дискретным ансамблем (X = { x, p(x) } - дискретный ансамбль)

Обозначим Ω = { A } - множество всевозможных подмножеств X
Тогда вероятность сложного события A:
$P\left(A\right)=\sum_{x\in A}p\left(x\right),\ A\in\Omega$
Т.е. вероятность сложного события равна сумме вероятностей простых событий, входящих в это сложное событие

Произведение событий A и B: $AB=A\cap B$
Для произвольной пары событий A, B ⊆ X (все элементы множеств A и B являются элементами множества X) условная вероятность (вероятность того, что произойдет событие A, если произошло событие B):
$P\left(A|B\right)=\dfrac{P\left(AB\right)}{P\left(B\right)},\ \textit{если}\ P\left(B\right)\neq0,\ \textit{иначе}\ 0$
Отсюда следует, что вероятность произведения событий $P\left(AB\right)=P\left(A|B\right)P\left(B\right)$
В общем случае:
$P\left(A_1\ldots A_n\right)=P\left(A_1\right)P\left(A_2|A_1\right)P\left(A_3|A_1A_2\right)\ldots P\left(A_n|A_1\ldots A_{n-1}\right)$

Если события A, B ∈ Ω независимы, то условная вероятность "превращается" в безусловную:
$P\left(A|B\right)=P\left(A\right);\ P\left(B|A\right)=P\left(B\right)$
$P\left(AB\right)=P\left(A\right)P\left(B\right)$
Соответственно, в общем случае при совместной независимости событий $A_1\ldots A_n$:
$P\left(A_1\ldots A_n\right)=P\left(A_1\right)P\left(A_2\right)\ldots P\left(A_n\right)$

Вероятность объединения событий:
$P\left(A\cup B\right)=P\left(A\right)+P\left(B\right)-P(AB)$



Вероятность объединения событий <= сумме вероятностей этих событий
$P\left(\bigcup_{m=1}^{M}A_m\right)\leq\sum_{m=1}^{M}P(A_m)$

Формула полной вероятности
Пусть даны M несовместных событий (т.е. их пересечение - пустое множество) $H_1,\ldots,H_M$ ("гипотез"), таких что вероятность их объединения = 1 ($P\left(\bigcup_{m=1}^{M}H_m\right)=1$). Тогда вероятность произвольного события A:
$P\left(A\right)=\sum_{m=1}^{M}P\left(A|H_m\right)P(H_m)$
Вспомним ранее выведенную формулу $P\left(AB\right)=P\left(A|B\right)P\left(B\right)$
Получим формулу апостериорной вероятности (формулу Байеса):
$P\left(H_j|A\right)=\dfrac{P(AH_j)}{P(A)}=\dfrac{P\left(A|H_j\right)P(H_j)}{\sum_{m=1}^{M}P\left(A|H_m\right)P(H_m)}$

Произведение ансамблей X = { x, p_X(x) } и Y = { y, p_Y (y) } определяется совместным распределением { p_{XY} (x, y) }, { (x, y) : x∈X, y∈Y }
Ансамбль произведений XY = { (x, y), p_{XY} (x, y) }

Условное распределение вероятностей:
$p\left(x|y\right)=\begin{cases}\frac{p(x,\,y)}{p(y)},\ \textit{если}\ p\left(y\right)\neq0\\ 0\ \textit{иначе}\end{cases},x\in X$

Ансамбли X и Y независимы, если $p\left(x,\ y\right)=p\left(x\right)p\left(y\right),\ x\in X,\ y\in Y$

Собственная информация - это количество информации, т.е. затраты (во времени или в пространстве), необходимые для передачи (хранения) данных

Данные	Представление
Числа (измерения)	Длина зависит от диапазона
Текст	8 бит (1 байт) на букву
Цифровая речь	13 бит на отсчет
Изображения (bmp)	3 байта на пиксель

Имеем:
X = { x, p(x) } - ансамбль
μ(x) - мера (количество) информации в x
Свойства количества информации:

1. Неотрицательность: μ(x) >= 0
2. μ(x) должна быть функцией от p(x) (чем больше вероятность, тем меньше информации несет сообщение, и наоборот)
3. Монотонность: если x, y ∈ X, p(x) >= p(y), тогда μ(x) <= μ(y)
4. Аддитивность: если x и y независимы, тогда μ(x, y) = μ(x) + μ(y) (если произошла серия событий, то их количество информации суммируется)
5. $\mu\left(p\left(x\right)^k\right) = k\mu\left(p\left(x\right)\right)$ (если есть серия одинаковых событий, то кол-во информации = кол-во событий * кол-во информации в одном событии)

Перечисленные требования приводят к следующему определению (формула собственной вероятности):
$I\left(x\right) = -log(p\left(x\right)),\ x \in X$

Чтобы узнать, насколько информативен ансамбль в целом, а не только отдельное входящее в него событие, нужно ввести понятие энтропии
Энтропия - это математическое ожидание от собственной информации
(Мат. ожидание - среднее значение случайной величины: $E\left(x\right) = \sum_{i=1}^{n} x_i p_i$ т.е. сумма произведений значений и их вероятностей)
$H\left(x\right) = E\left[I\left(x\right)\right] = \sum_x p\left(x\right) I\left(x\right) = \sum_x -p\left(x\right) log(p(x))$
Энтропия показывает, насколько можно максимально сжать информацию, т.е. предел сжатия, измеряется в битах на символ
Свойства энтропии:

1. Неотрицательность H(X) >= 0
2. H(X) <= log|X| (равенство, если все элементы X равновероятны) (|X| - алфавит)
3. Если X = { x, p(x) } и Y = { y = f(x), p(y) }, тогда H(Y) <= H(X) (с равенством, если f обратима)
4. Если X и Y независимы, тогда H(XY) = H(X) + H(Y)
5. H(X) - выпуклая вверх функция распределения вероятностей на элементах ансамбля X
6. Пусть X = { x, p(x) } и A ⊆ X (A - подмножество X). Введем ансамбль X' = { x, p'(x) } и p'(x) как:
$$p'\left(x\right) = \begin{cases} \frac{P(A)}{|A|},\ x \in A, \\ p\left(x\right),\ x \notin A. \end{cases}$$
(т.е. часть вероятностей усреднили, см. первое уравнение)
Тогда H(X') >= H(X)
7. Если для двух ансамблей X и Y распределения вероятностей отличаются только порядком следования элементов, то H(X) = H(Y)

Доказательство свойства 2
$$\begin{aligned} H(X) - \log|X| &\overset{(a)}{=} -\sum_{x\in X} p(x)\log p(x) - \sum_{x\in X} p(x)\log|X| = \\ &\overset{(b)}{=} \sum_{x\in X} p(x)\log\frac{1}{p(x)|X|} \le \\ &\overset{(c)}{\le} \log e\left[\sum_{x\in X} p(x)\left(\frac{1}{p(x)|X|} - 1\right)\right] = \\ &= \log e\left(\sum_{x\in X}\frac{1}{|X|} - \sum_{x\in X} p(x)\right) = 0 \quad . \end{aligned}$$
Доказательство (c):
$\ln x \le x - 1 \iff \log x \le (x-1)\log e.$

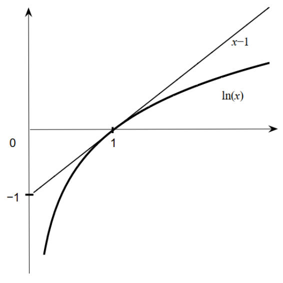


Рис.: Графическая интерпретация $\ln(x) \le x - 1$

Доказательство свойства 5
Энтропия H(p) ансамбля с распределением вероятностей p - выпуклая вверх функция от p. Докажем это.
Запишем формулу энтропии:
$H\left(p\right) = \sum_{m=1}^{M} -p_m log(p_m)$
Есть свойство выпуклых функций, что сумма выпуклых функций выпукла.
Докажем, что функция $f_m = -p_m log(p_m)$ выпукла. Для этого найдем вторую производную:
$f_m'' = -log(e)/p_m$
Для любой вероятности $p_m \in (0, 1)$ вторая производная отрицательна => функция выпуклая вверх

Частный случай: двоичный ансамбль

- $X = \{0, 1\}$. Пусть $p(1) = p,\ p(0) = 1 - p = q.$
- Энтропия двоичного ансамбля

$\triangleq$ - равно по определению

- Энтропия двоичного ансамбля

$$H(X) = -p\log p - q\log q \triangleq \eta(p).$$

по определению (просто обозначение такое есть)

- Первая производная от $\eta(p)$.

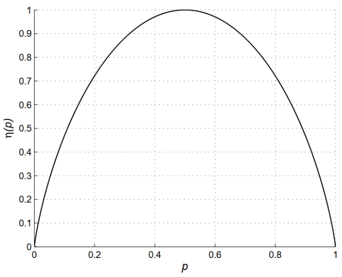
$$\eta'(p) = -\log p + \log(1-p),$$

$\eta'(p)$=0, при $p = \frac{1}{2}$ – точка экстремума.

- Вторая производная от $\eta(p)$.

$$\eta''(p) = -\log e/p - \log e/(1-p) < 0.$$

$$H(X) = \eta(p) = -p\log p - (1-p)\log(1-p)$$



Доказательство свойства 6

- Обозначим $\tilde{\boldsymbol{p}} = ((p_1+p_2)/2, (p_1+p_2)/2, p_3, ..., p_M)$.
- Необходимо доказать, что

$$H(\tilde{\boldsymbol{p}}) \geq H(\boldsymbol{p}).$$

- Обозначим

$$\begin{aligned} \boldsymbol{p}' = \boldsymbol{p} &= (p_1, p_2, p_3, ..., p_M), \\ \boldsymbol{p}'' &= (p_2, p_1, p_3, ..., p_M). \end{aligned}$$

т.к. если вероятности те же, хоть и в другом порядке, энтропия та же

- Заметим, что: $H(\boldsymbol{p}') = H(\boldsymbol{p}'') = H(\boldsymbol{p})$.
- $\tilde{\boldsymbol{p}} = (\boldsymbol{p}' + \boldsymbol{p}'')/2$.
- Из выпуклости энтропии следует, что:

$$H(\tilde{\boldsymbol{p}}) = H\left(\frac{\boldsymbol{p}' + \boldsymbol{p}''}{2}\right) \geq \frac{1}{2}H(\boldsymbol{p}') + \frac{1}{2}H(\boldsymbol{p}'') = H(\boldsymbol{p}).$$

т.к. энтропия – выпуклая функция, функция от среднего ≥ среднему от функции