Условная собственная информация. Условная энтропия и её свойства

4 января 2023 г. 1:52

(Из вопроса "Дискретный ансамбль". Начало)
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:

$$p\left(x|y\right)=\begin{cases}\frac{p(x,\,y)}{p(y)},\; если\; p\left(y\right)\neq 0\\ 0\; иначе\end{cases},\; x\in X$$

Ансамбли X и Y независимы, если $p\left(x,\ y\right)=p\left(x\right)p\left(y\right),\ x\in X,\ y\in Y$
(Из вопроса "Дискретный ансамбль". Конец)

Собственная информация - это количество информации, т.е. затраты (во времени или 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: $E\left(x\right)=\sum_{i=1}^{n}x_ip_i$ 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

Условная собственная информация сообщения x при фиксированном y:
$I\left(x|y\right)=-log(p\left(x|y\right))$
Условная энтропия X при заданном y∈Y:
$H\left(X|y\right)=-\sum_{x\in X}p\left(x|y\right)log(p(x|y))$
Условная энтропия X при фиксированном ансамбле Y:
$H\left(X|Y\right)=-\sum_{y\in Y}\left(p\left(y\right)\sum_{x\in X}p\left(x|y\right)log(p(x|y))\right)=-\sum_{x\in X}\sum_{y\in Y}p\left(x,y\right)log(p(x|y))$
(пояснение: $p\left(x|y\right)=\dfrac{p(x,\ y)}{p(y)}$, $p(x,\ y)$ - вероятность пары x и y)

Свойства условной энтропии:

1. H(X|Y) >= 0
2. H(X|Y) <= H(X), равенство, если X и Y независимы
3. H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y) - это энтропия для пары X и Y (типа блок из двух символов)
4. H(X|YZ) <= H(X|Y), равенство, если X и Z условно независимы для всех y∈Y
5. $H\left(X_1\ldots X_n\right)=H\left(X_1\right)+H\left(X_2|X_1\right)+H\left(X_3|X_1X_2\right)+\ldots+H\left(X_n|X_1,\ldots,X_{n-1}\right)$
6. $H\left(X_1\ldots X_n\right)\le\sum_{i=1}^{n}H\left(X_i\right)$, равенство, если $X_1,\ldots,X_n$ являются совместно независимыми

Доказательство свойства 2 для условной энтропии.

$$p(x)=\sum_{y\in Y}p(x|y)p(y)$$

$$\begin{aligned}H(X|Y)-H(X) &= -\sum_{x\in X}\sum_{y\in Y}p(x,y)\log p(x|y)+\\ &+\sum_{x\in X}\sum_{y\in Y}p(x,y)\log p(x)=\\ &=\sum\sum p(x,y)\log\frac{p(x)}{p(x|y)}\le\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{x\in X}\sum_{y\in Y} p(x,y)\left(\frac{p(x)}{p(x|y)}-1\right)\log e = \\
&= \left(\sum_{x\in X}\sum_{y\in Y} p(y)p(x) - \sum_{x\in X}\sum_{y\in Y} p(x,y)\right)\log e = \\
&= 0.
\end{aligned}
$$

Доказательство свойства 3 для энтропии.

- Рассмотрим $X = \{x, p(x)\}$, $f(x)$, $Y = \{y = f(x), x \in X\}$.
- Нужно доказать, что

$$H(Y) \leq H(X).$$

- Используя свойство 3 условной энтропии:

$$H(XY) = \underbrace{H(X|Y)}_{\geq 0} + H(Y) = \underbrace{H(Y|X)}_{=0} + H(X).$$

- Поскольку $f(x)$ определена для каждого x, получим, что $H(Y|X) = 0$, $H(X|Y) \geq 0$.