Часть 6

ОСНОВЫ КЛАССИЧЕСКОЙ ТЕОРИИ ИНФОРМАЦИИ

Классическая энтропия. Биты и наты

Классические компьютеры работают в двоичной логике "1" и "0" (есть напряжение в ячейке или нет). В этом случае считают, что каждая ячейка содержит один бит информации. Регистр из N ячеек, очевидно, содержит N бит информации. С их помощью можно записать 2^N сообщений. Если сообщения равновероятны, то вероятность каждого сообщения $w_N = 1/2^N$. Введем информационную энтропию по формуле: $H = -\log_2 w_N$. Для нашего простейшего случая $H = N$. На самом деле, есть сообщения более вероятные и менее вероятные. Поэтому Клод Шеннон обобщил информационную энтропию следующим образом:

$$H = -\sum_N w_N \ln(w_N).$$

Последняя формула записана через натуральный логарифм, а не через $\log_2$. Поэтому информацию по Шеннону измеряют в "натах":

$$1\,\textbf{бит} = 1\,\textbf{нат}/\ln 2 \approx 1,44\,\textbf{нат}.$$

Энтропия и информация

Вернемся к записи сообщений в регистр из N ячеек. Если все сообщения равновероятны, то до прочтения содержания регистра мы не имеем никакой информации о том, какое из 2^N сообщений в нем записано. Пусть I – мера информации. Тогда до прочтения регистра $I_{in} = 0$, но энтропия $H_{in} = N$. Когда сообщение из регистра прочитано, то мы стали обладать информацией в N бит, то есть $I_{fin} = N$. При этом мы точно знаем, какое сообщение записано в регистре. Поэтому $w_{N\,fin} = 1 \Rightarrow H_{fin} = 0$. Таком образом, в нашем примере выполняется следующее равенство

$$H_{in} + I_{in} = N = H_{fin} + I_{fin}.$$

В общем случае, связь энтропии и информации задается следующим образом:

$$H + I = const$$

или

$$\Delta H = -\Delta I.$$

Следовательно, шенноновскую информационную энтропию можно рассматривать как меру недостатка информации о классической системе. Чем меньше энтропия, тем больше информация. В этом случае говорят о том, что система упорядочена. Чем больше энтропия, тем меньше информации о системе. Часто говорят о том, что система приближается к состоянию хаоса.

Декогеренция и парадокс кота Шредингера

С понятием энтропии тесно связан вопрос о переходе суперпозиции состояний в смесь состояний.

Пусть живой котик помещен в непрозрачный ящик с атомом в возбужденном состоянии. У атома (для простоты) имеются всего две возможности: 1) остаться в возбужденном состоянии; 2) испустить фотон и перейти в основное состояние. Предположим, что испущенный фотон всегда попадает в гранату, которая, непременно, взрывается после поглощения фотона, и, естественно, убивает котика своими осколками.

Пусть $\hat{\rho}_1$ – матрица плотности системы, когда атом не излучал, граната не взорвалась и котик жив. А $\hat{\rho}_2$ – матрица плотности для случая, когда атом излучил фотон, граната взорвалась и котик погиб ради идеалов абстрактной науки.

Из раздела "Суперпозиция или смесь!" ясно, что матрица плотности системы "атом + граната + котик" описывается либо смесью, либо суперпозицией матриц плотности $\hat{\rho}_1$ и $\hat{\rho}_2$.

Если в любой момент времени экспериментатор Аленушка откроет непрозрачный ящик, то она увидит либо живого, либо (печалька!) мертвого котика, но никогда живо-мертвого или мертво-живого. То есть, Аленушка всегда увидит смесь макроскопически различных состояний, но никогда не увидит их суперпозицию. Вопрос: почему так происходит, ведь ясно, что атом находился в суперпозиции состояний? Где и как произошла декогеренция и суперпозиция превратилась в смесь?

Короткий ответ на поставленные вопросы: **во всем виновато наше... Солнце!**

В более развернутой форме. Все предметы на Земле, в том числе и макроприборы, находятся в потоке излучения Солнца, который у поверхности Земли составляет $\frac{\Delta Q}{\Delta t} \approx 1,4 \cdot 10^6$ **эрг/см^2 сек** (так называемая солнечная постоянная). Этот поток излучения, в конечном счете, преобразуется в тепловую энергию атомов и молекул со средней температурой $< T > \sim 300$ **K**. Тогда поток изменения энтропии от Солнца (**в битах!**) у поверхности Земли равен:

$$\frac{\Delta H_\odot}{\Delta t} \sim -\frac{1}{\Delta t}\frac{\Delta Q}{k_B < T >} = -\frac{1}{k_B < T >}\frac{\Delta Q}{\Delta t} \approx -3 \cdot 10^{19}\,\frac{\text{бит}}{\text{см}^2\,\text{сек}},$$

где k_B – постоянная Больцмана. Потоку солнечной энтропии должен соответствовать поток информации

$$\frac{\Delta I_\odot}{\Delta t} = -\frac{\Delta H_\odot}{\Delta t} \approx 3 \cdot 10^{19}\,\frac{\text{бит}}{\text{см}^2\,\text{сек}}.$$

В этом потоке информации "купаются" все макроприборы. Взаимодействие с потоком информации от Солнца приводит к **постоянному измерению** состояния макроприбора внешней средой и, следовательно, к редукции матрицы плотности макроприбора. Но редукция матрицы плотности макроприбора автоматически приводит к редукции запутанной (в смысле модели измерения фон Неймана) матрицы плотности **"микросистема + макроприбор"** (в рассматриваемом случае это матрица плотности $\hat{\rho}$ системы **"атом + граната + котик"**).

Оценим время Δt_R между актами редукции. Чтобы произошла редукция, необходимо передать информацию о состоянии макроприбора. Минимальное количество требуемой информации – 1 **бит**. Поэтому:

$$1\,\text{бит} \sim \frac{\Delta I_{\odot}}{\Delta t}\,\Delta t_R\, L^2,$$

где L – характерный размер макроприбора. Примем $L = 10$ **см**. Тогда

$$\Delta t_R \sim 3 \cdot 10^{-22}\ \text{сек}.$$

Это время необходимо сравнить с характерным временем срабатывания макроприбора. Даже для лучших макроприборов оно не превосходит $\Delta t_D \sim 10^{-10}$ **сек**.

Поэтому, пока макроприбор производит одно измерение над микросистемой, в объединенной системе происходит порядка $\Delta t_D/\Delta t_R \approx 10^{11}$ **актов редукции**. Столько же раз меняется относительная фаза между состояниями $\hat{\rho}_1$ и $\hat{\rho}_2$! Следовательно, процесс измерения должен включать в себя усреднение по огромному числу этих различных относительных фаз. Основываясь на данной оценке и результатах раздела "Суперпозиция или смесь!" приходим к выводу, что в земных условиях появление мертво-живых и живо-мертвых котиков маловероятно. Этим и разрешается парадокс кота Шредингера.

Парадокс был предложен Э.Шредингером в 1935 году в процессе полемики вокруг другого парадокса – парадокса Эйнштейна-Подольского-Розена – и, по вполне понятным причинам, получил название **парадокса кота Шредингера**.

Граница между мирами

Исходя из представленной выше модели декогеренции мы способны оценить масштаб, начиная с которого систему уже можно считать микроскопической и применять к ней квантовые законы.

Согласно соотношению неопределенностей, $\Delta p \Delta x \sim \hbar$. Положим $\Delta x \sim L$, а характерное изменение импульса оценим как:

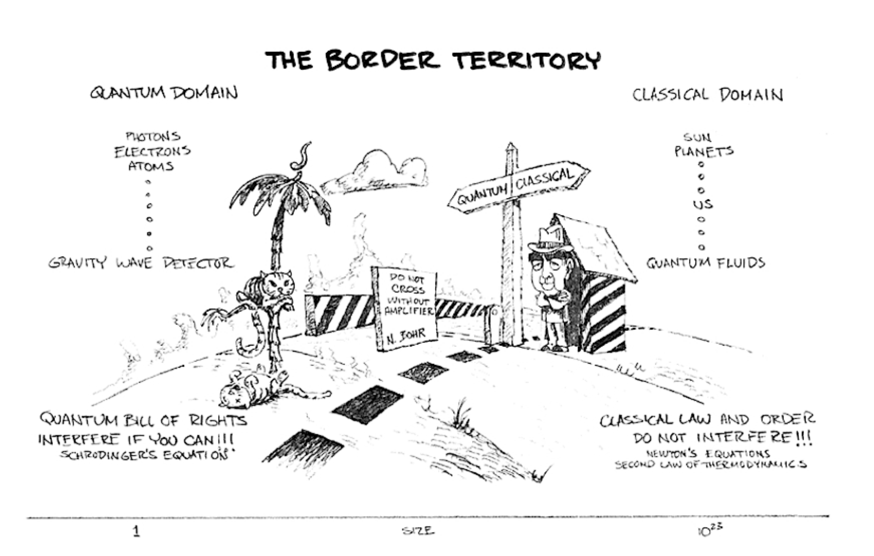
$$\Delta p \sim mv \sim m\frac{L}{\Delta t_R} \sim \rho\, L^3\, L\, \frac{\Delta I_\odot}{\Delta t}\, L^2 = \rho\, \frac{\Delta I_\odot}{\Delta t}\, L^6,$$

гдн ρ – характерная плотность тела. Таким образом:

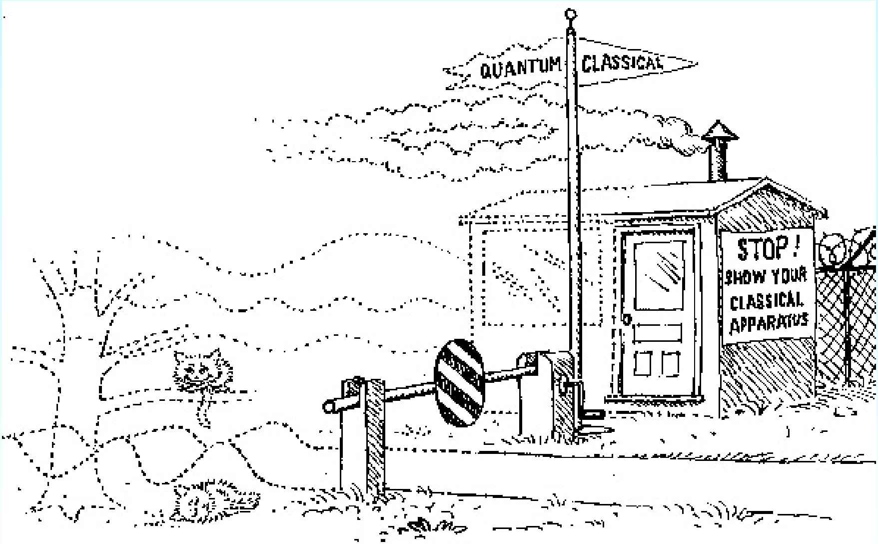
$$\frac{\rho}{\hbar}\,\frac{\Delta I_\odot}{\Delta t}\, L^7 \sim 1 \quad \Rightarrow \quad L \sim \left(\frac{\hbar}{\rho}\,\frac{1}{\Delta I_\odot/\Delta t}\right)^{1/7}.$$

Если принять $\rho \sim 10$ гр/см^3, то $L \sim 10^{-7}$ см $\ll 10^{-4}$ см – длины волны видимого света. Таким образом, квантовые явления нельзя видеть невооруженным глазом. Заметим, что L порядка размеров типичных молекул.

А так “на самом деле” выглядит граница между квантовым и классическим мирами по мнению знакомого нам W.Zurek-a – одного из авторов “No-cloning theorem”.



Или так, пока наблюдатель на границе спит...



Ограничения на величину шенноновской энтропии

Введенную выше энтропии по Шеннону $H = -\sum\limits_N w_N \ \ln{(w_N)}$ можно определить более строго (как это принято в теории информации). Пусть имеется набор из N величин x_ℓ, каждая из которых может появляться с вероятностью $1 \geq w(x_\ell) \geq 0$ (например, буквы в тексте, грани несимметричной игральной кости и т.д.). Тогда говорят, что задан ансамбль $X = \{x_\ell,\, w(x_\ell)\}$, шенноновская энтропия которого есть

$$H(X) \;=\; -\sum_{\ell=1}^{N} w(x_\ell)\, \ln{(w(x_\ell))}, \quad \textbf{где} \quad \sum_{\ell=1}^{N} w(x_\ell) \;=\; 1.$$

Поскольку все $w(x_\ell)$ неотрицательны и не превосходят единицы, то очевидно, что $H(X) \geq 0$. Равенство достигается, когда $w(x_k) = 1$ и $w(x_1) = \ldots = w(x_{k-1}) = w(x_{k+1}) = \ldots = w(x_N) = 0$.

Теперь найдем максимальное значение энтропии $H(X)$. Из условия нормировки вероятности имеем:

$$w(x_N) = 1 - \sum_{\ell=1}^{N-1} w(x_\ell).$$

Тогда энтропию можно записать в виде

$$H(X) = -\sum_{\ell=1}^{N-1} w(x_\ell)\ \ln\left(w(x_\ell)\right) - w(x_N)\ \ln\left(w(x_N)\right).$$

При $k \neq N$ с учетом условия нормировки получаем

$$\frac{\partial}{\partial w(x_k)}\Big(w(x_N)\ \ln\left(w(x_N)\right)\Big) = -\ \ln\left(w(x_N)\right) - 1.$$

Максимум $H(X)$ находим из условия

$$0 = \frac{\partial H(X)}{\partial w(x_k)} = \ln\left(w(x_N)\right) - \ln\left(w(x_k)\right).$$

Таким образом, $w(x_k) = w(x_N)$. Поскольку k пробегает любые целые значения от 1 до $N-1$, то максимум функции $H(X)$ достигается, когда вероятности $w(x_1) = w(x_2) = \ldots = w(x_{N-1}) = w(x_N) = 1/N$. И этот максимум равен $H_{max}(X) = -N\,\frac{1}{N}\ \ln\left(\frac{1}{N}\right) = \ln\left(N\right)$.

Из вышесказанного следует, что энтропия Шеннона $H(X)$ лежит в диапазоне

$$0 \leq H(X) \leq \ln\left(N\right).$$

где N - количество элементов в ансамбле X.

Двоичная энтропия

В теории информации (как классической, так и квантовой) важную роль играет ансамбль X_2, который состоит всего из двух элементов x_1 и x_2, каждый из которых возникает с вероятностью $W(x_1) = p$ и $W(x_2) = 1 - p$ соответственно ($0 \leq p \leq 1$). Энтропия такого ансамбля

$$H(X_2) \equiv H(p) = -p \ln p - (1-p) \ln (1-p)$$

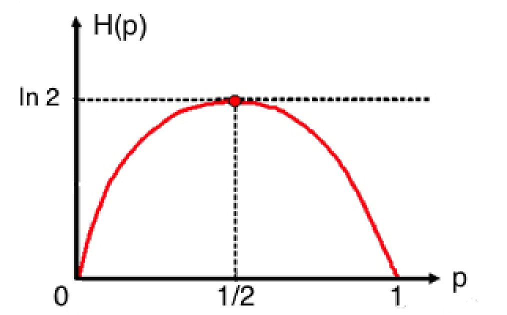
носит название двоичной энтропии.

Очевидно, что двоичная энтропия обладает симметрией относительно значения $p = 1/2$, то есть

$$H(p) = H(1-p).$$

Легко проверить, что максимальное значение двоичной энтропии $H_{max} = \ln 2$ достигается при $p = 1/2$, минимальное – $H_{min} = 0$ – при $p = 0$ и $p = 1$ в полном согласии с общими ограничениями, полученными на предыдущей странице.

График функции $H(p)$ ведет себя следующим образом:



Из графика видно, что двоичная энтропия является **вогнутой функцией** (то есть любая прямая, соединяющая две точки на графике лежит ниже графика самой функции $H(p)$). Отсюда сразу получаем, что

$$H(px_1 + (1-p)x_2) \geq pH(x_1) + (1-p)H(x_2).$$

Равенство достигается только если $x_1 = x_2$ или $p = 0$, или $p = 1$. Данное свойство называется вогнутостью двоичной энтропии.

Неравенство Йенсена и вогнутость энтропии Шеннона

Свойство вогнутости можно сформулировать математически более строго. Функция $f(x)$ называется вогнутой при $x \in [a, b]$, если $f''_{xx} \leq 0$.

Для вогнутой функции выполняется следующее неравенство, которое носит название неравенства Йенсена. Пусть функция $f(x)$ вогнутая на интервале $[a, b]$ и пусть точки x_1, ..., x_n принадлежат этому интервалу. Кроме того, пусть w_1, ..., w_n – некоторые положительные числа, удовлетворяющие условию $\sum\limits_{\ell=1}^{n} w_\ell = 1$. Тогда:

$$f\left(\sum_{\ell=1}^{n} w_\ell x_\ell\right) \geq \sum_{\ell=1}^{n} w_\ell f\left(x_\ell\right).$$

Элементарное доказательство этого факта можно найти, например, на стр.10 книги Ю. П. Соловьёва "Неравенства", Издательство Московского центра непрерывного математического образования, Москва, 2005 г.

Наконец, в качестве функции $f(x)$ рассмотрим функцию $f(x) = -x \ln x$. Легко проверить, что эта функция является вогнутой при $x \geq 0$. Тогда неравенство Йенсена приводит к свойству вогнутости для энтропии Шеннона:

$$H\left(\sum_{\ell=1}^{n} w_\ell x_\ell\right) \geq \sum_{\ell=1}^{n} w_\ell H\left(x_\ell\right).$$

Классическая относительная энтропия и неравенство Гиббса

Пусть теперь для величин x_ℓ существуют два различных набора вероятностей $P = \{x_\ell,\, p(x_\ell)\}$ и $Q = \{x_\ell,\, q(x_\ell)\}$, где $0 \le p(x_\ell) \le 1$, $0 \le q(x_\ell) \le 1$ и $\sum\limits_{\ell=1}^{N} p(x_\ell) = \sum\limits_{\ell=1}^{N} q(x_\ell) = 1$. Тогда классической относительной энтропией распределения "P" относительно распределения "Q" (relative entropy) называется величина

$$H\Big(P\,||\,Q\Big) = -\sum_{\ell=1}^{N} p(x_\ell)\,\ln q(x_\ell) - H(P) = \sum_{\ell=1}^{N} p(x_\ell)\,\ln\left(\frac{p(x_\ell)}{q(x_\ell)}\right).$$

С аналогичной величиной мы уже встречались в разделе "Количественное сравнение квантовых состояний". И там эта величина носила название метрики Кульбака-Лейблера.

Поэтому $H\Big(P\,||\,Q\Big)$ нужно рассматривать как классическую меру различия двух вероятностных распределений одного и того же набора величин $\{x_\ell\}$.

В дальнейших вычислениях всегда будем полагать, что $0 \ln 0 = 0$ и $-p(x_\ell) \ln 0 = +\infty$.

Теперь докажем, что классическая относительная энтропия неотрицательна, т.е. что

$$H\Big(P \,||\, Q\Big) \geq 0.$$

Это неравенство называется неравенством Гиббса. Доказательство основано на простом факте, что при $x \geq 0$ верно неравенство $-\ln x \geq 1 - x$ (проверьте это самостоятельно при помощи разложения в ряд Тейлора). Тогда:

$$\begin{aligned} H\Big(P \,||\, Q\Big) &= -\sum_{\ell=1}^{N} p(x_\ell) \ln\left(\frac{q(x_\ell)}{p(x_\ell)}\right) \geq \sum_{\ell=1}^{N} p(x_\ell) \left(1 - \frac{q(x_\ell)}{p(x_\ell)}\right) = \\ &= \sum_{\ell=1}^{N} p(x_\ell) - \sum_{\ell=1}^{N} q(x_\ell) = 1 - 1 = 0, \quad \text{ч.т.д.} \end{aligned}$$

Равенство достигается, когда оба распределения совпадают, т.е. когда $p(x_\ell) = q(x_\ell)$. Неравенство Гиббса удобно использовать для исследования свойств других энтропийных величин.

Классическая совместная энтропия и субаддитивность

Пусть имеется два набора случайных величин $X = \{x_\ell\}$ и $Y = \{y_m\}$, для которых определены совместные вероятности $w(x_\ell, y_m)$. Тогда для этих наборов естественно определить понятие классической совместной энтропии (joint entropy)

$$H(X, Y) = -\sum_{\ell, m} w(x_\ell, y_m) \ln w(x_\ell, y_m).$$

Поскольку $w(x_\ell, y_m) = w(y_m, x_\ell)$ и $\sum\limits_{\ell, m} \ldots = \sum\limits_{m, \ell} \ldots$, то

$$H(X, Y) = H(Y, X).$$

Совместная энтропия является мерой полной неопределенности для пары наборов случайных величин (X, Y). Она обладает свойством субаддитивности (доказательство будет дано позже)

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

Равенство достигается, когда наборы X и Y являются независимыми, т.е. когда $w(x_\ell, y_m) = w(x_\ell)\, w(y_m)$.

Очевидно, что понятие совместной энтропии может быть расширено на любое число наборов случайных величин.

Условная вероятность и теорема Байеса

Обозначим через $w\left(y_m|x_\ell\right)$ – условную вероятность найти величину y_m при условии, что величина x_ℓ уже известна. Тогда по теореме Байеса

$$w\left(y_m|x_\ell\right) = \frac{w(y_m,\, x_\ell)}{w(x_\ell)} = \frac{w(x_\ell,\, y_m)}{w(x_\ell)}.$$

При этом

$$w(y_m) = \sum_\ell w(y_m,\, x_\ell) = \sum_\ell w\left(y_m|x_\ell\right)\, w(x_\ell)$$

и

$$w(x_\ell) = \sum_m w(x_\ell,\, y_m) = \sum_m w\left(x_\ell|y_m\right)\, w(y_m).$$

Из теоремы Байеса сразу следует, что

$$w\left(x_\ell|y_m\right)\, w(y_m) = w(x_\ell,\, y_m) = w(y_m,\, x_\ell) = w\left(y_m|x_\ell\right)\, w(x_\ell).$$

Классическая условная энтропия

Распишем выражение для совместной энтропии с учетом теоремы Байеса. Имеем:

$$
\begin{aligned}
H(X,\,Y) &= -\sum_{\ell,m} w(x_\ell, y_m) \ln w(x_\ell, y_m) = \\
&= -\sum_{\ell,m} w(x_\ell, y_m) \ln \Big(w\,(y_m|x_\ell)\, w(x_\ell) \Big) = \\
&= -\sum_{\ell,m} w(x_\ell, y_m) \ln w\,(y_m|x_\ell) - \sum_{\ell,m} w(x_\ell, y_m) \ln w(x_\ell) \; = \\
&= -\sum_{\ell,m} w(y_m, x_\ell) \ln w\,(y_m|x_\ell) - \sum_{\ell} w(x_\ell) \ln w(x_\ell) \; = \\
&= H\,(Y|X) + H(X).
\end{aligned}
$$

Величина

$$
\begin{aligned}
H\,(Y|X) &= -\sum_{\ell,m} w(y_m, x_\ell) \ln w\,(y_m|x_\ell) = \\
&= -\sum_{\ell} w(x_\ell) \sum_{m} w\,(y_m|x_\ell) \ln w\,(y_m|x_\ell) = \sum_{\ell} w(x_\ell)\, H\,(Y|x_\ell)
\end{aligned}
$$

называется классической условной энтропией (conditional entropy) или общей классической условной энтропией.

А величина

$$H\left(Y|x_\ell\right) = -\sum_m w\left(y_m|x_\ell\right) \ln w\left(y_m|x_\ell\right)$$

носит название частной классической условной энтропии.

Функция $H\left(Y|X\right)$ служит мерой неопределенности ансамбля Y при известном значении ансамбля X. А функция $H\left(Y|x_\ell\right)$ служит мерой неопределенности ансамбля Y при известном значении случайной величины x_ℓ.

Поскольку $0 \leq w\left(y_m|x_\ell\right) \leq 1$, то $-\ln w\left(y_m|x_\ell\right) \geq 0$. Поэтому

$$H\left(Y|X\right) \geq 0 \quad \text{и} \quad H\left(Y|x_\ell\right) \geq 0.$$

Аналогично величине $H\left(Y|X\right)$ можно определить классическую условную энтропию $H\left(X|Y\right)$ из соотношения:

$$H\left(X|Y\right) = H(X,\ Y) - H(Y).$$

Очевидно, что $H\left(X|Y\right) \geq 0$.

Доказательство субаддитивности

Выше было получено, что $H(Y|X) \geq 0$ и $H(X|Y) \geq 0$. Отсюда немедленно следуют неравенства:

$$H(X,\, Y) \geq H(X) \text{ и } H(X,\, Y) \geq H(Y).$$

Теперь исполним обещание и докажем субаддитивность энтропии. Для этого воспользуемся неравенством Гиббса. Рассмотрим наборы $P = \{x_\ell,\, y_m,\, w(x_\ell,\, y_m)\}$ и $Q = \{x_\ell,\, y_m,\, w(x_\ell)\, w(y_m)\}$. Тогда:

$$\begin{aligned}
0 \leq H\Big(P||Q\Big) &= -\sum_{\ell,\, m} w(x_\ell,\, y_m) \ln \left(\frac{w(x_\ell)\, w(y_m)}{w(x_\ell,\, y_m)}\right) = \\
&= -\sum_{\ell} \ln w(x_\ell) \sum_{m} w(x_\ell,\, y_m) - \sum_{m} \ln w(y_m) \sum_{\ell} w(x_\ell,\, y_m) + \\
&\quad + \sum_{\ell,\, m} w(x_\ell,\, y_m) \ln w(x_\ell,\, y_m) = -\sum_{\ell} w(x_\ell) \ln w(x_\ell) - \\
&\quad - \sum_{m} w(y_m) \ln w(y_m) - H(X, Y) = H(X) + H(Y) - H(X, Y).
\end{aligned}$$

Неравенство доказано. Окончательно получаем, что

$$2 \cdot H(X,\, Y) \geq H(X) + H(Y) \geq H(X,\, Y).$$

Классическая взаимная информация

Классической взаимной информацией (**mutual information**) называется информация, которая является общей для наборов X и Y. Она записывается в виде:

$$I(X:Y) = H(X) + H(Y) - H(X,Y).$$

откуда немедленно следует, что $I(X:Y) \geq 0$ **(см. неравенство в конце предыдущего слайда). По своему определению** $I(X:Y)$ **должна быть симметричной функцией наборов** X **и** Y**, то есть:**

$$I(X:Y) = I(Y:X) \geq 0.$$

Отметим полезное равенство

$$I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

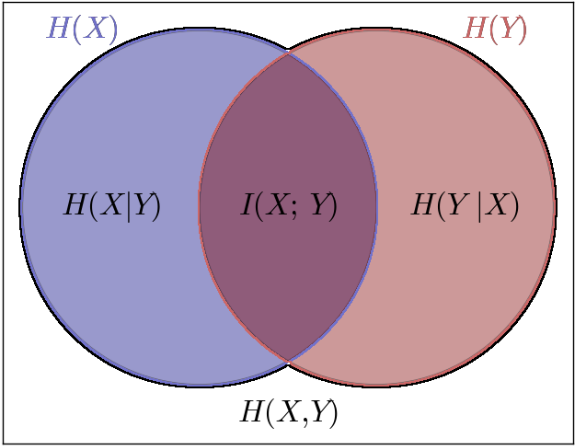
Из этого равенства и неотрицательности взаимной информации сразу получаем, что

$$H(X) \geq H(X|Y) \quad \text{и} \quad H(Y) \geq H(Y|X).$$

Заметим, что взаимная информация НЕ обладает свойством субаддитивности, т.е.

$$I(X,Y:Z) \nleq I(X:Z) + I(Y:Z).$$

Наглядная связь между различными энтропиями и взаимной информацией



Сильная субаддитивность

Довольно очевидно, что можно узнать больше информации о наборе X, если известна информация о наборах Y и Z, которые каким то образом связанны с набором X (например, при помощи совместных распределений вероятностей или найденных на опыте эмпирических корреляций), чем когда известна информация только об одном таком наборе Y. В силу соотношения между энтропией и информацией для условной энтропии должно иметь место обратное неравенство

$$H(X|Y,Z) \le H(X|Y).$$

Распишем правую и левую части данного неравенств согласно определению условной энтропии. Имеем

$$H(X|Y,Z) = H(X,Y,Z) - H(Y,Z) \quad \text{и} \quad H(X|Y) = H(X,Y) - H(Y).$$

Отсюда немедленно получаем свойство сильной субаддитивности для энтропии Шеннона

$$H(X,Y,Z) + H(Y) \le H(H,Y) + H(Y,Z).$$