

# ЭНТРОПИЯ

Марина Аюшеева, Яна Коротова, Олеся Майстренко, Елизавета Махнева, Дарья Писарева

## Что это такое?

Вспомним всем известную игру «данетки». Так, чтобы понять, о чем идет речь, мы задаем уточняющие вопросы. Так вот, минимальное число вопросов, необходимое, чтобы выяснить полную информацию об объекте, является *энтропией*.

В теории информации энтропия – *степень неопределенности, связанная со случайной величиной*.<sup>1</sup>

Также энтропию можно определить как *наименьшее среднее число бит, необходимое для кодирования некоторой информации*:

$$H = - \sum_{i=1}^n p_i \log p_i$$

где  $p_i$  — вероятность  $i$ -го исхода. Или вероятность того, что в «данетках» угадываемый объект обладает некоторой характеристикой. Например, нужно угадать, какого человека загадали, и с вероятностью  $2/3$  он моложе 30 лет, с вероятностью  $1/3$  старше 30 лет. Такое может быть в ситуации, когда молодых людей среди тех, кого могли бы загадать, больше, или загадывающий отдает предпочтение более молодым людям.

**Задача 1.** Красная Шапочка встретила соседа-лесоруба Николая Петровича по дороге к бабушке Елене, которая равновероятно может жить в одной из трех деревень (Шапка точно помнит, в какой именно). Поскольку девочка маленькая (а неподалеку обитает волк), тот решил узнать, в какой деревне живет бабушка, только не спросив напрямую, а задавая наводящие вопросы.

Найдите энтропию местонахождения Елены.

## Еще немножко :)

**Условная энтропия** — количество бит, необходимое для того, чтобы узнать имеющуюся информацию о случайной величине  $Y$  при условии, что случайная величина  $X$  известна.

Можно объяснить и проще — вспомним вновь игру выше. Вам необходимо узнать, кого загадал человек, ведущий в «данетке». Однако теперь он загадывает не одного человека, а *пару*. Каждый человек в этой паре с равными вероятностями может быть как моложе 30 лет (в двух случаях из трех), так и старше 30 (в одном случае из трех). И нам известно, что точно загадали человека моложе 30 лет (одному человеку из этой пары меньше 30 лет). Это и будет наше условие  $X$ . А далее мы уже исходя из данной информации должны отгадать, кого же все-таки загадали?

Энтропия (среднее число вопросов, необходимое для того, чтобы понять, кого загадали) рассчитывается так:

$$H(Y|X) = - \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)}$$

<sup>1</sup><https://stackoverflow.com/questions/510412/what-is-the-computer-science-definition-of-entropy>

**Задача 2.** Оказалось, на дороге в одну из трёх деревень (в каждой из которых равновероятно может находиться бабуля Елена) ошивается злой волк Матвей, а в одну деревню ведет только одна дорога. Вероятности того, что Матвей находится в деревне  $i$ -той ( $X$  — местонахождение волка по вертикали), и того, что Елена в деревне  $j$ -той ( $Y$  — местонахождение бабули по горизонтали):

$$\begin{pmatrix} 1/4 & 1/24 & 1/24 \\ 1/24 & 1/4 & 1/24 \\ 1/24 & 1/24 & 1/4 \end{pmatrix}$$

Найдите условную энтропию местонахождения Матвея:  $H(X|Y)$ .

**Совместная энтропия** — степень неопределенности, связанная со множеством случайных величин.

Как и ранее, ведущий загадал пару людей. Однако теперь мы ничего заранее не знаем, кроме вероятностей, с которыми могли загадать людей, обладающих определенными признаками. Иными словами, вероятность, с которой загадали человека моложе 30, вероятность, с которой волосы загаданного человека имеют рыжий оттенок, и так далее.

Совместная энтропия (среднее число вопросов, необходимое для отгадывания пары людей) рассчитывается так:

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log p(x, y)$$

**Задача 3.** Оказалось, на дороге в одну из трёх деревень (в каждой из которых равновероятно может находиться бабуля Елена) ошивается злой волк Матвей, а в одну деревню ведет только одна дорога. Вероятности того, что Матвей находится в деревне  $i$ -той ( $X$  — местонахождение волка по вертикали), и того, что Елена в деревне  $j$ -той ( $Y$  — местонахождение бабули по горизонтали):

$$\begin{pmatrix} 1/4 & 1/24 & 1/24 \\ 1/24 & 1/4 & 1/24 \\ 1/24 & 1/24 & 1/4 \end{pmatrix}$$

Найдите совместную энтропию местонахождения Елены и Матвея:  $H(X, Y)$ .

**Все упомянутые выше герои обладают следующими свойствами:**

◇  $H \geq 0$

◇  $H(Y|X) = H(X, Y) - H(X)$  или в более общем случае  $H(X_1, \dots, X_n) = \sum_{i=1}^n H(X_i | X_1, \dots, X_n)$

◇  $H(Y|X) \leq H(Y)$

◇  $H(X, Y) = H(X|Y) + H(Y|X) + I(X; Y) = H(X) + H(Y) - I(X; Y)$ , где  $I(X; Y)$  — взаимная информация о случайных величинах  $X$  и  $Y$

◇  $I(X; Y) \leq H(X)$ ,  $I(X; Y) = H(X) - H(X|Y)$

**Взаимная информация**  $I(X; Y)$  — мера взаимной зависимости двух случайных величин.

Другими словами, **взаимная информация** — это те сведения, которые мы получаем, отгадав одного из людей в паре, когда играем в «данетки». Так, чтобы отгадать одного из людей в паре ( $X$ ), в среднем необходимо  $H(X)$  вопросов. Чтобы отгадать этого человека, однако зная, что второй загаданный это  $Y$ , в среднем необходимо  $H(X|Y)$  вопросов. Тогда разница в количестве вопросов и есть **взаимная информация**, или те вопросы, которые уже не нужно задавать, если известен один из людей в паре. Также совместная информация обладает свойством *симметричности*:  $I(X; Y) = I(Y; X)$ . Интуитивно это объясняется тем, что если нам известен один человек из загаданной пары, то часть вопросов (число которых в среднем равно  $H(X)$ ) задавать уже не имеет смысла, поскольку ответ на них следует из уже полученной информации об известном человеке.

Рассчитывается она так:

$$I(X; Y) = H(X) - H(X|Y) = - \sum_{x \in X} p(x) \log p(x) + \sum_{x \in X, y \in Y} p(x, y) \log \frac{p(x, y)}{p(y)} = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

**Задача 4.** Оказалось, на дороге в одну из трёх деревень (в каждой из которых равновероятно может находиться бравый охотник Борис, защищающий бабулю Елену от диких зверей) ошивается злой волк Матвей, а в одну деревню ведет только одна дорога. Вероятности того, что Матвей находится в деревне  $i$ -той ( $X$  — местонахождение волка по вертикали), и того, что Борис в деревне  $j$ -той ( $Y$  — местонахождение бабули по горизонтали):

$$\begin{pmatrix} 1/5 & 1/15 & 1/15 \\ 1/15 & 1/5 & 1/15 \\ 1/15 & 1/15 & 1/5 \end{pmatrix}$$

Найдите взаимную информацию местонахождения Матвея и Бориса:  $I(X; Y)$

## А есть еще кросс энтропия!

**Кросс энтропия** — минимальное среднее количество бит, необходимое для того, чтобы закодировать некоторую информацию, если схема кодирования базируется на некотором распределении  $q$ , а не истинном,  $p$ .

$$CH(p, q) = - \sum_{i=1}^n p_i \log q_i$$

Также кросс энтропию можно определить через *расстояние Кульбака – Лейблера*. Для начала стоит узнать, что это:

**Расстояние Кульбака – Лейблера** — степень отдаленности друг от друга двух вероятностных распределений (называется также *относительная энтропия*).

Интуитивная интерпретация данного понятия такова: насколько больше вопросов мы зададим в случае неоптимальной стратегии игры в «данетки», чем в случае оптимальной стратегии. Можно заметить, что расстояние Кульбака – Лейблера равно разности кросс-энтропии и

энтропии, так как на кодирование информации при неоптимальной стратегии уйдет гораздо больше бит, чем при оптимальной. Мы ищем именно ту разницу, насколько неоптимальная стратегия хуже оптимальной:

$$D(P \parallel Q) = CH(p, q) - H(p), \text{ или } CH(p, q) = H(p) + D_{KL}(p \parallel q)$$

Рассчитывается для дискретного случая так:

$$D(P \parallel Q) = - \sum_{i=1}^n p_i \log q_i - (- \sum_{i=1}^n p_i \log p_i)$$

**Задача 5.** Красная Шапочка, убегая от злого лесоруба Николая Петровича, в панике перепутала вероятности, с которыми охотник Борис находится в одной из деревень ( $X$  — местонахождение охотника):

$$(1/6 \quad 2/3 \quad 1/6),$$

и с которыми волк Матвей ошивается на одной из дорог в деревни ( $Y$  — местонахождение волка):

$$(3/8 \quad 3/8 \quad 1/4).$$

- (а) Найдите кросс-энтропию местонахождения Матвея;
- (б) Вычислите дивергенцию Кульбака-Лейблера.

## А что, только для дискретных случайных величин?

Нет! :)

В случае, если Вы работаете с абсолютно непрерывными случайными величинами, энтропия и её родственники рассчитываются по следующим формулам:

- ◇ Самая главная и простая энтропийка:

$$H(X) = - \int_{-\infty}^{+\infty} f(x) \log f(x) dx$$

- ◇ Условная энтропия:

$$H(Y|X) = - \int_{-\infty}^{+\infty} f(x, y) \log f_{Y|X}(y) dy$$

- ◇ Совместная энтропия:

$$H(X, Y) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \log f(x, y) dx dy$$

- ◇ Взаимная информация:

$$I(X; Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \log \frac{f(x, y)}{f(x)f(y)} dx dy$$

◇ Кросс-энтропия:

$$CH(p, q) = - \int_{-\infty}^{+\infty} p(x) \log q(x) dx$$

◇ Расстояние Кульбака – Лейблера:

$$D(P || Q) = \int_{-\infty}^{+\infty} f(x) \log f(x) dx - \int_{-\infty}^{+\infty} f(x) \log g(x) dx$$

**Задача 6.** На дороге в деревню, где живет бабушка Елена, равномерно разбросаны монетки. Дорога в эту деревню располагается на отрезке  $[0; A]$ . Красная Шапочка решила закодировать эту информацию и отправить бабуле с помощью драконой почты. Сколько бит ей понадобится для этого при оптимальной схеме кодирования?

**Задача 7.** Злой лесоруб Николай Петрович решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку Елену, которая против их отношений. Лесоруб не знает, чем конкретно занята сейчас бабушка, но он знает, что она может делать.

Бабушка Елена пробегает от крайней левой деревни до крайней правой (от точки  $A = -1$  до точки  $B = 1$ ), причём её местоположение отлично описывается следующим распределением:

$$f_{\xi}(x) = \begin{cases} 1 - \frac{3}{4}x^2, & x \in [-1; 1] \\ 0, & \text{иначе} \end{cases}$$

После того, как Елена пробегает все расстояние, она бежит к кусту (её местоположение равномерно меняется от точки 1 до кустика с ягодами). Если местоположение куста с ягодами является случайной величиной, имеющей равномерное распределение на отрезке  $[-1; 1]$ , то какова условная энтропия местоположения бабушки?

**Задача 8.** Злой лесоруб Николай Петрович решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку Елену, которая против их отношений. Лесоруб не знает, чем конкретно занята сейчас бабушка, но он знает, что она может делать.

Бабушка Елена пробегает от крайней левой деревни до крайней правой (от точки  $A = -1$  до точки  $B = 1$ ), причём её местоположение отлично описывается следующим распределением:

$$f_{\xi}(x) = \begin{cases} 1 - \frac{3}{4}x^2, & x \in [-1; 1] \\ 0, & \text{иначе} \end{cases}$$

После того, как Елена пробегает все расстояние, она бежит к кусту (её местоположение равномерно меняется от точки 1 до кустика с ягодами). Если местоположение куста с ягодами является случайной величиной, имеющей равномерное распределение на отрезке  $[-1; 1]$ , то какова совместная энтропия местонахождения Елены и куста с ягодами?

**Задача 9.** Злой лесоруб Николай Петрович решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку Елену, которая против их отношений. Лесоруб не знает, чем конкретно занята сейчас бабушка, но он знает, что она может делать.

Бабушка Елена пробегает от крайней левой деревни до крайней правой (от точки  $A = -1$  до точки  $B = 1$ ), причём её местоположение отлично описывается следующим распределением:

$$f_{\xi}(x) = \begin{cases} 1 - \frac{3}{4}x^2, & x \in [-1; 1] \\ 0, & \text{иначе} \end{cases}$$

После того, как Елена пробегает все расстояние, она бежит к кусту (её местоположение равномерно меняется от точки 1 до кустика с ягодами). Если местоположение куста с ягодами является случайной величиной, имеющей равномерное распределение на отрезке  $[-1; 1]$ , то какова взаимная информация местоположения бабушки и куста с ягодами?

**Задача 10.** Красная Шапочка должна отнести бабуле пирожки (не только же ягодами ей питаться!). Пока Красная Шапочка бежала к бабушке Елене из её корзинки в какой-то момент начали выпадать пирожки. Всего в корзинке их было  $n$  штук. Известно, что пирожки выпадали равномерно на некотором отрезке дороги. Потерю всех пирожков Шапка обнаружила лишь по прибытии к бабушке и сразу решила собрать все выпавшие пирожки. Предполагая, что расстояние от дома Шапочки до дома Бабушки равно  $a$ , рассчитайте кросс-энтропию, если:

- (а) Шапочка знает, что пирожки выпадали равномерно;
- (б) Шапочка не знает, что пирожки выпадали равномерно;
- (в) Какая из величин больше?

## Чуть-чуть истории...

В 1948 году, исследуя проблему рациональной передачи информации через зашумлённый коммуникационный канал, **Клод Шеннон** предложил революционный вероятностный подход к пониманию коммуникаций и создал первую, истинно математическую, теорию энтропии.

Его сенсационные идеи быстро послужили основой разработки двух основных направлений: *теории информации*, которая использует понятие вероятности для изучения статистических характеристик данных и коммуникационных систем, и *теории кодирования*, в которой используются главным образом алгебраические и геометрические инструменты для разработки эффективных кодов.

Понятие энтропии, как меры случайности, введено Шенноном в его статье «*Математическая теория связи*» (англ. A Mathematical Theory of Communication), опубликованной в двух частях в Bell System Technical Journal в 1948 году.

В случае равновероятных событий (частный случай), остается зависимость только от количества рассматриваемых вариантов, и формула Шеннона значительно упрощается и совпадает с *формулой Хартли*, которая впервые была предложена американским инженером **Ральфом Хартли в 1928 году**, как один из научных подходов к оценке сообщений:

$$I = -\log p = \log N,$$

где  $I$  — количество передаваемой информации,  $p$  — вероятность события,  $N$  — возможное количество различных (равновероятных) сообщений.

## Применение энтропии и ее родственников

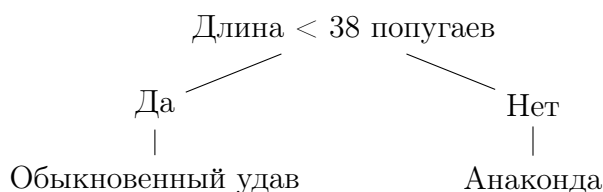
## Энтропийное кодирование

Как говорилось ранее, энтропия показывает наименьшее среднее число бит, необходимое для кодирования некоторой информации. Данное свойство используется, как ни странно, при кодировании информации.

Например, код Шеннона-Фано. С целью минимизации энтропии и, соответственно, оптимизации кода элементы с большой вероятностью появления кодируются меньшим числом символом. Таким образом, производится сжатие объема информации, что позволяет передавать большее количество информации, затрачивая меньший объем памяти.

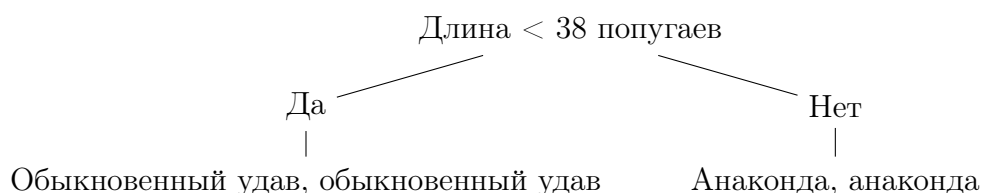
## Построение решающих деревьев

Решающие деревья — метод, использующийся в машинном обучении и работающий по принципу принятия решений человеком. Каждое ветвление представляет собой разделение выборки на две части по порогу некоторого признака. Например, признак — длина, пороговое значение — 38. Все объекты, длина которых превышает 38, отделяются от объектов с длиной меньше 38 и дальнейший анализ проходят отдельно.



В данном методе расчет энтропии помогает определить оптимальный порог для каждого узла решения. А именно, подбирается такое разделение выборки, при котором взвешенная сумма энтропий получившихся выборок минимальна среди возможных вариантов разбиений.

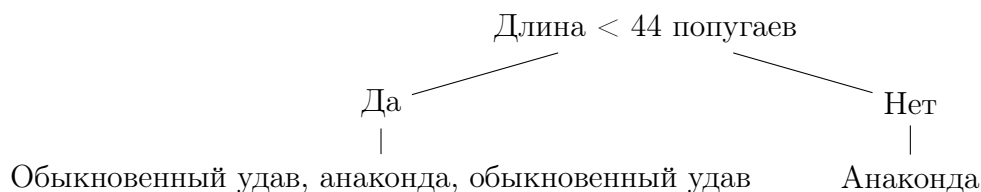
Например, у нас есть выборка объектов с одним признаком, длина: обыкновенный удав (22 попугая), анаконда (46 попугаев), анаконда (40 попугаев), обыкновенный удав (31 попугай). Мы выбираем порог: 38 или 44 попугаев? Попробуем разделить выборку по 38 попугаям:



При расчете энтропии  $0 \cdot \log_2 0$  считается равным 0, несмотря на  $\log_2 0$ . За вероятность принимается вероятность встретить данный класс в новой выборке.

Энтропия левой части:  $-(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = 0$ . Энтропия правой части:  $-(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = 0$ . Суммарная энтропия получилась:  $\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 0 = 0, \frac{1}{2}$  — доля каждой выборки в исходной.

Попробуем разделить выборку по 44 попугаям:



Энтропия левой части:  $-(\frac{1}{3} \cdot \log_2 \frac{1}{3} + \frac{2}{3} \cdot \log_2 \frac{2}{3}) \approx 0.92$ . Энтропия правой части:  $-(1 \cdot \log_2 1 + 0 \cdot \log_2 0) = 0$ . Суммарная энтропия получилась:  $\frac{3}{4} \cdot 0.92 + \frac{1}{4} \cdot 0 = 0.69$ .

В первом случае мы идеально разделили выборку при энтропии, равной нулю. Во втором случае нам удалось отделить одну анаконду, но не удалось отделить классы. Так и энтропия в первом случае оказалась меньше, чем во втором. Причем ее равенство нулю необязательно — любое значение меньше 0.69 показало бы, что первый случай более оптимален. Здесь же, в виду неотрицательности энтропии, однозначно можно сказать, что критерий «длина < 38 попугаев» дает оптимальный результат.

Энтропия позволяет получать наименее разнообразные по содержанию классов выборки. Соответственно, признак и пороговое значение подбираются наиболее оптимально — алгоритм успешно отделяет объекты, принадлежащие к одному классу.

## Применение в алгоритмах *t-SNE* и *UMAP*

В анализе данных часто возникает необходимость в снижении размерности, и в таких случаях на помощь приходят знания об энтропии. Речь, конечно, идет не об энтропии как таковой, а об алгоритмах, которые базируются на теории.

При создании пространства меньшей размерности, *t-SNE* и *UMAP* используют кросс-энтропию как показатель эффективности перенесения свойств объектов. Чем меньше кросс-энтропия, тем ближе к истинному оказалось подобранное отображение.

Приведем пример работы алгоритма *UMAP*. Мы возьмем набор данных об одежде, который включает в себя 70000 черно-белых изображений различной одежды по 10 классам: футболки, брюки, свитеры, платья, кроссовки и т.д. Каждая картинка имеет размер 28x28 пикселей или 784 пикселя.

Изначально каждый пиксель является признаком объекта (фотографии) и принимает некоторое значение (цвет). Если бы каждая картинка состояла из двух пикселей, мы бы смогли построить график, где по оси абсцисс отложен цвет одного пикселя, по оси ординат — цвет второго пикселя, и изобразить все объекты точками.

У нас каждая картинка состоит из 784 пикселей — 784-мерное пространство, поэтому изобразить его проблематично. Но если мы преобразуем выборку таким образом, что останется всего два признака, то мы сможем визуализировать ее. Получившиеся признаки будут уже не пикселями, а абстрактными признаками, которые алгоритм получает из исходных — переводит 784 признака в 2 с помощью функции, подобранной в результате работы.

Реализуем описанный алгоритм<sup>2</sup>.

Внимание! Библиотека *UMAP* требует предварительной установки<sup>3</sup>.

Импортируем нужные библиотеки:

```
import numpy as np # работа с матрицами
from mnist import MNIST # наборы данных
import matplotlib.pyplot as plt # построение графиков
%matplotlib inline
import umap # алгоритм UMAP
```

Загружаем набор данных с фотографиями одежды.

```
mndata = MNIST('fashionmnist')
train, train_labels = mndata.load_training()
```

<sup>2</sup>Реализация позаимствована из <https://github.com/lmcinnes/umap/blob/master/doc/supervised.rst>

<sup>3</sup>Почитать про установку: <https://umap-learn.readthedocs.io/en/latest/>



```
test, test_labels = mndata.load_testing()
data = np.array(np.vstack([train, test]), dtype=np.float64) / 255.0
target = np.hstack([train_labels, test_labels])
```

Создаем список из наименований одежды.

```
classes = ['T-shirt/top', 'Trouser', 'Pullover', 'Dress', 'Coat', 'Sandal',
'Shirt', 'Sneaker', 'Bag', 'Ankle boot']
```

Запускаем UMAP.

```
embedding = umap.UMAP().fit_transform(data)
```

Визуализируя результат, получаем:

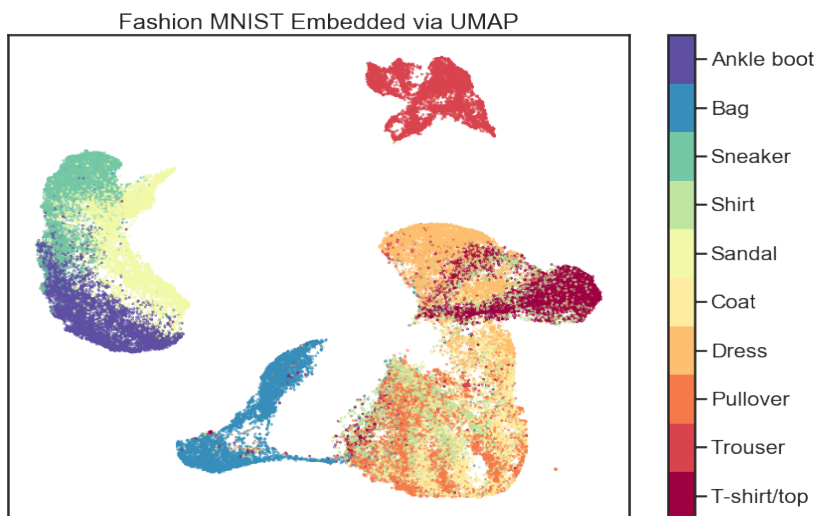


Рис. 1: Алгоритм UMAP

Получить два новых признака из исходных можно очень многими способами. Но они должны описывать выборку как можно лучше, чтобы при визуализации мы видели не случайно нарисованное изображение, а отображение начального пространства.

Именно тут пригодится кросс-энтропия, но начнем немного издалека. В UMAP используется дивергенция Кульбака-Лейблера для случайной величины Бернулли  $X \sim B(p(x))$ . Здесь  $p(x)$  — некоторая функция, определяющая вероятность того, что  $X = 1$ . Запишем формулу  $D_{KL}$  для  $X$ :

$$\begin{aligned} D_{KL}(P \parallel \tilde{P}) &= -p(x) \log \tilde{p}(x) - (1 - p(x)) \log(1 - \tilde{p}(x)) + p(x) \log p(x) + (1 - p(x)) \log(1 - p(x)) = \\ &= p(x) \log \frac{p(x)}{\tilde{p}(x)} + (1 - p(x)) \log \frac{1 - p(x)}{1 - \tilde{p}(x)} \end{aligned}$$

Однако алгоритм рассчитывает не просто разницу между двумя распределениями для одной случайной величины, а сумму таких разниц для  $n$  случайных величин:

$$C_P(\tilde{P}) = \sum_{i=1}^n p(x_i) \log \frac{p(x_i)}{\tilde{p}(x_i)} + (1 - p(x_i)) \log \frac{1 - p(x_i)}{1 - \tilde{p}(x_i)}$$

Данная величина показывает степень отдаленности друг от друга множеств из случайных величин:  $P$  и  $\tilde{P}$ . При этом каждому множеству соответствует своя функция:  $p(x)$  и  $\tilde{p}(x)$ .

Минимизация  $C_P(\tilde{P})$  по  $\tilde{p}(x)$  позволяет найти множество  $\tilde{P}$ , которое наиболее похоже на множество  $P$ .

Задача минимизации заключается в поиске оптимального  $\tilde{p}(x)$ . Если вернуться к исходной записи  $D_{KL} = H_p(\tilde{p}) - H(p)$ , то видно, что энтропия не зависит от  $\tilde{p}(x)$ , соответственно, является константой при минимизации. Тогда задача преобразуется в оптимизацию лишь суммы кросс-энтропий:

$$-\sum_{i=1}^n (p(x_i) \log \tilde{p}(x_i) + (1 - p(x_i)) \log(1 - \tilde{p}(x_i))) \rightarrow \min_{\tilde{p}}$$

Поэтому UMAP считается методом, основанным на кросс-энтропии.

При отображении пространстве UMAP строит взвешенный граф из объектов. Веса ребер можно воспринимать как вероятность существования данного ребра. Поэтому ребро  $e$  является случайной величиной, имеющей распределение Бернулли:  $e \sim B(w(e))$ , где  $w(e)$  — вес ребра  $e$ . Получается, что множество ребер построенного графа — множество  $E$  из случайных величин Бернулли.

Тогда для более корректного переноса данных мы можем подобрать для множества  $E_h$  похожее на него множество  $E_l$  с функцией  $w_l(e)$ , соответствующие низкоразмерному пространству.

Например, пусть у нас есть выборка из 3 объектов, и мы знаем веса ребер между ними в исходном пространстве:

	D	Y	L
D	-	0.7	0.1
Y	0.7	-	0.3
L	0.1	0.3	-

UMAP выполняет автоматическую минимизацию кросс-энтропии. Поскольку вручную данное действие выполнить сложно, то мы можем попробовать подобрать  $t(m)$  с наименьшей дивергенцией Кульбака-Лейблера, выбирая из нескольких множеств:

$t_1$	D	Y	L
D	-	0.6	0.3
Y	0.6	-	0.5
L	0.3	0.5	-

$t_2$	D	Y	L
D	-	0.1	0.9
Y	0.1	-	0.4
L	0.9	0.4	-

Кросс-энтропия для  $t_1(m)$ :

$$\begin{aligned} C_{E_h}(E_l)_1 &= 0.7 \log \frac{0.7}{0.6} + (1 - 0.7) \log \frac{1 - 0.7}{1 - 0.6} + 0.1 \log \frac{0.1}{0.3} + (1 - 0.1) \log \frac{1 - 0.1}{1 - 0.3} + \\ &+ 0.3 \log \frac{0.3}{0.5} + (1 - 0.3) \log \frac{1 - 0.3}{1 - 0.5} \approx 0.22 \end{aligned}$$

Кросс-энтропия для  $t_2(m)$ :

$$\begin{aligned} C_{E_h}(E_l)_2 &= 0.7 \log \frac{0.7}{0.1} + (1 - 0.7) \log \frac{1 - 0.7}{1 - 0.1} + 0.1 \log \frac{0.1}{0.9} + (1 - 0.1) \log \frac{1 - 0.1}{1 - 0.9} + \\ &+ 0.3 \log \frac{0.3}{0.4} + (1 - 0.3) \log \frac{1 - 0.3}{1 - 0.4} \approx 2.81 \end{aligned}$$

Так как наша цель — минимальная дивергенция Кульбака-Лейблера, то множество весов с функцией  $t_1(m)$  подходит для решения задачи больше, чем  $t_2$ . Вероятно, можно подобрать еще более оптимальное множество, однако проще доверить эту работу UMAP.

## Ключи к сердцу Красной Шапочки

Тут можно найти ответы и решения ко всем задачам, которые были представлены выше. Переходите к этому разделу, только если уже всё решили и хотите проверить себя :)

**Задача 1.** Теперь подсчитаем количество возможных вариантов размещения трех агентов: Шапки, бабули и волка. Общее число размещений каждого из трех агентов в трех местах равно 27. Далее, чтобы узнать вероятность хорошего исхода, необходимо знать, во скольких случаях бабушка не получает свои пирожки (то есть Шапка и волк оказываются в одной деревне вместе). Они могут оказаться в одной деревне только вдвоём или с бабулей, но и том, и в другом случае сказка не имеет счастливого конца. Таких вариантов девять. Тогда вариантов, в которых бабуля наслаждается пирожками 18. Энтропия равна:

$$H = - \sum_{i=1}^{18} \frac{1}{18} \log \frac{1}{18}$$

Результатом подсчёта данного выражения будет число 4.17. Это означает, что в среднем (при большом количестве повторений эксперимента) ей будет достаточно 4.17 вопросов, чтобы угадать деревню и доставить пирожки. При округлении до ближайшего целочисленного дает значение пять. Это означает, что в среднем за пять вопросов возможно узнать, получила ли бабушка свои пирожки и в какой деревне она находится, а вот за четыре вопроса это будет сделать сложнее.

Ответ: а) 18 вариантов; б) 4.17

**Задача 2.** Всего вариантов шесть (три деревни и Волка не было, волк был, но его поймали). Так как все три деревни равновероятны, рассмотрим одну из них. Вероятность того, что волка не было:

$$\frac{3 - N}{3}$$

Вероятность того, что он был и его поймали:

$$\frac{N}{3} \cdot \frac{1}{3}$$

Одно из событий наступило гарантированно:

$$\frac{3 - N}{3} + \frac{N}{9} = \frac{9 - 2N}{9}$$

Тогда вероятность  $\mathbb{P}$ (Волк был в деревне А и его поймали):

$$\frac{N}{9} \cdot \frac{9}{(9 - 2N) \cdot 3} = \frac{N}{(9 - 2N) \cdot 3}$$

$\mathbb{P}$ (Волка не было в деревне А):

$$\frac{(3 - N)}{3} \cdot \frac{9}{(9 - 2N) \cdot 3} = \frac{3 - N}{9 - 2N}$$

Аналогично для других деревень.

Получаем энтропию:

$$H = 3 \cdot \frac{N}{(9 - 2N) \cdot 3} \cdot \log \frac{(9 - 2N) \cdot 3}{N} + 3 \cdot \frac{3 - N}{9 - 2N} \cdot \log \frac{9 - 2N}{3 - N}$$

Ответ: см. в решении

**Задача 3.** Волк равновероятно находится в любой из деревень: в А, Б и В. Вероятность каждого из них равна  $1/3$ . Бабушка живет в деревнях с разной вероятностью. Таким образом вероятность того, что они были в деревне А равна  $1/6$ , так как мы перемножили  $1/2$  и  $1/3$ . Аналогично, вероятность того, что они были в деревнях Б и В равна  $1/12$ .

Итого совместная вероятность:

$$H = - \left( \frac{1}{6} \log \frac{1}{6} + \frac{1}{12} \log \frac{1}{12} + \frac{1}{12} \log \frac{1}{12} \right)$$

Ответ:  $\frac{1}{2} + \frac{1}{3} \cdot \log 3$

**Задача 4.** Сюрприз! Вам предоставляется невероятная возможность самостоятельно проверить свои силы и решить такую задачу. Варианты решения мы просим присылать в issues на открытый репозиторий на сайте [github.com](https://github.com)<sup>4</sup>. Первое правильное решение поощряется специальным ценным призом.

Желаем удачи и с нетерпением ждем Ваших решений!

**Задача 5.** Съесть пирожок с капустой можно 20 вариантами (а шапочка считает, что 10). Итого, 10 пирожков с капустой и КШ считает, что они с капустой, а 10 с капустой и КШ не считает их с капустой. Тогда:

$$H = 10 \cdot \frac{1}{20} \cdot \log 10 = \frac{\log 10}{2}$$

А решение второго пункта мы также принимаем в issues на открытый репозиторий [github.com](https://github.com). Первый правильно решивший задачу получит ценный приз от терверисток из группы БЭК171! :)

Ответ: а)  $\frac{\log 10}{2}$ , б) ждем Ваших решений!

**Задача 6.**  $X$  — расстояние, которое прошла Шапка:

$$f(x) = \frac{1}{B}$$

$$H(x) = - \int_0^B \frac{1}{B} \log_2 \left( \frac{1}{B} \right) dx = \frac{1}{B} \cdot \log_2 B \cdot B = \log_2 B$$

Ответ:  $\log_2 B$

**Задача 7.**  $Y$  — расстояние, которое прошла Шапка:

$$f(x, y) = \frac{1}{B} \cdot \frac{1}{x} \Rightarrow f_{Y|X}(y) = \frac{1}{B}$$

$$H(Y|X) = - \int_0^B \frac{1}{B} \cdot \frac{1}{x} \cdot \log_2 \left( \frac{1}{B} \right) dy = \frac{1}{x} \log_2 B$$

Ответ:  $\frac{\log_2 B}{x}$

---

<sup>4</sup>Ссылка на репозиторий: <https://github.com/oomaystrenko/entropy>