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, среднее число вопросов, необходимое, чтобы выяснить полную информацию об объекте, является *энтропией*.

В теории информации энтропия – *степень неопределенности, связанная со случайной величиной.*[1]

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 по дороге к бабушке, которая равновероятно может жить в одной из трех деревень. Шапка точно помнит, в какой именно. Поскольку девочка маленькая, а неподалеку обитает волк, лесоруб решил узнать, в какой деревне живет бабушка, только не спросив напрямую, а задавая наводящие вопросы.

Найдите энтропию местонахождения бабули.

Еще немножко :)

> **Условная энтропия** — количество бит, необходимое для того, чтобы узнать значение случайной величины 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?

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

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

[1] 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, 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 бабушка в деревне j-той (Y — местонахождение бабули по горизонтали):

Деревня	2112	2110	К10
2112	$1/4$	$1/24$	$1/24$
2110	$1/24$	$1/4$	$1/24$
К10	$1/24$	$1/24$	$1/4$

Найдите условную энтропию местонахождения Матвея при неизвестном местонахождении бабули: $H(X|Y)$.

> **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:

$$H(X,Y) = -\sum_x \sum_y p(x,y) \log p(x,y)$$

Задача 3. Оказалось, на дороге в одну из трёх деревень, в каждой из которых равновероятно может находиться бабуля, 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 бабушка в деревне j-той (Y — местонахождение бабули по горизонтали):

Деревня	2112	2110	К10
2112	$1/4$	$1/24$	$1/24$
2110	$1/24$	$1/4$	$1/24$
К10	$1/24$	$1/24$	$1/4$

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

> **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:

$$I(X;Y) = H(X) - H(X|Y) =$$

$$= -\sum_x p(x)\log p(x) + \sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(y)} = \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

Задача 4. Оказалось, на дороге в одну из трёх деревень, в каждой из которых равновероятно может находиться бабуля, 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 бабушка в деревне j-той (Y — местонахождение бабули по горизонтали):

Деревня	2112	2110	К10
2112	$1/4$	$1/24$	$1/24$
2110	$1/24$	$1/4$	$1/24$
К10	$1/24$	$1/24$	$1/4$

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

Все упомянутые выше герои обладают следующими свойствами:

- ◇ $H \geqslant 0$
- ◇ $H(Y|X) = H(X,Y) - H(X)$
- ◇ $H(Y|X) \leqslant H(Y)$
- ◇ $H(X,Y) = H(X|Y) + H(Y|X) + I(X;Y) = H(X) + H(Y) - I(X;Y)$, где $I(X;Y)$ – взаимная информация о случайных величинах X и Y
- ◇ $I(X;Y) \leqslant H(X),\ \ I(X;Y) = H(X) - H(X|Y)$

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.

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

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

> **Дивергенция Кульбака – Лейблера** — степень отдаленности одного вероятностного распределения от другого (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 дивергенция Кульбака – Лейблера равна разности кросс-энтропии и энтропии, так как на кодирование информации при неоптимальной стратегии уйдет гораздно больше бит, чем при оптимальной. Мы ищем именно ту разницу, насколько неоптимальная стратегия хуже оптимальной:

$$D_{KL}(P\,||\,Q) = CE(P||Q) - H(p), \text{ или } CE(P||Q) = H(p) + D_{KL}(P\,||\,Q)$$

Рассчитывается для дискретного случая так:

$$D_{KL}(P\,||\,Q) = -\sum_{i=1}^{n} p_i \log q_i \; - (-\sum_{i=1}^{n} p_i \log p_i)$$

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

$$\begin{pmatrix} 1/6 & 2/3 & 1/6 \end{pmatrix},$$

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

$$\begin{pmatrix} 3/8 & 3/8 & 1/4 \end{pmatrix}.$$

(а) Найдите кросс-энтропию из истинного распределения местонахождения Матвея в распределение местонахождения Бориса;

(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 расчитываются 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:

$$I(X;Y) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy$$

◇ Кросс-энтропия:

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

◇ Дивергенция Кульбака – Лейблера:

$$D_{KL}(P\,||\,Q) = \int\limits_{-\infty}^{+\infty} p(x)\log p(x)dx - \int\limits_{-\infty}^{+\infty} p(x)\log q(x)dx$$

Задача 6. На дороге в деревню, где живет бабушка, Красная Шапка обронила серебрянную монетку, местонахождение которой распределено равномерно. Дорога в эту деревню располагается на отрезке $[0; A]$. Шапка решила сообщить бабуле энтропию местонахождения потерянной монетки. Помогите бедной Шапочке её посчитать.

Задача 7. Злой лесоруб решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку, которая против их отношений. Лесоруб не знает, где именно находится бабушка.

Бабушка ест ягодки. Местоположение кустика с ягодками X и местоположение ямы Y, которую выкопала Красная Шапочка для деревца, отлично описываются многомерным нормальным распределением:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\right)$$

Какова условная энтропия местоположения бабушки в зависимости от местоположения ямы?

Задача 8. Злой лесоруб решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку, которая против их отношений. Лесоруб не знает, где именно находится бабушка.

Бабушка ест ягодки. Местоположение кустика с ягодками X и местоположение ямы Y, которую выкопала Красная Шапочка для деревца, отлично описываются многомерным нормальным распределением:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\right)$$

Какова совместная энтропия местоположения бабушки и местоположения ямы?

Задача 9. Злой лесоруб решил, что он должен завладеть сердцем Красной Шапки и устранить со своего пути её бабушку, которая против их отношений. Лесоруб не знает, где именно находится бабушка.

Бабушка ест ягодки. Местоположение кустика с ягодками X и местоположение ямы Y, которую выкопала Красная Шапочка для деревца, отлично описываются многомерным нормальным распределением:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}\right)$$

Какова взаимная информация местоположения бабушки и местоположения ямы?

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 символов. 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.

Длина $<$ 38 попугаев

Да — Обыкновенный удав

Нет — Анаконда

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:

Длина $<$ 38 попугаев

Да — Обыкновенный удав, обыкновенный удав

Нет — Анаконда, анаконда

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:

Длина $<$ 44 попугаев

Да — Обыкновенный удав, анаконда, обыкновенный удав

Нет — Анаконда

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 алгоритме 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, UMAP использует 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[2].

Внимание! Библиотека UMAP требует предварительной установки[3].

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()
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:

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

[2] Реализация позаимствована из `https://github.com/lmcinnes/umap/blob/master/doc/supervised.rst`

[3] Почитать про установку: `https://umap-learn.readthedocs.io/en/latest/`

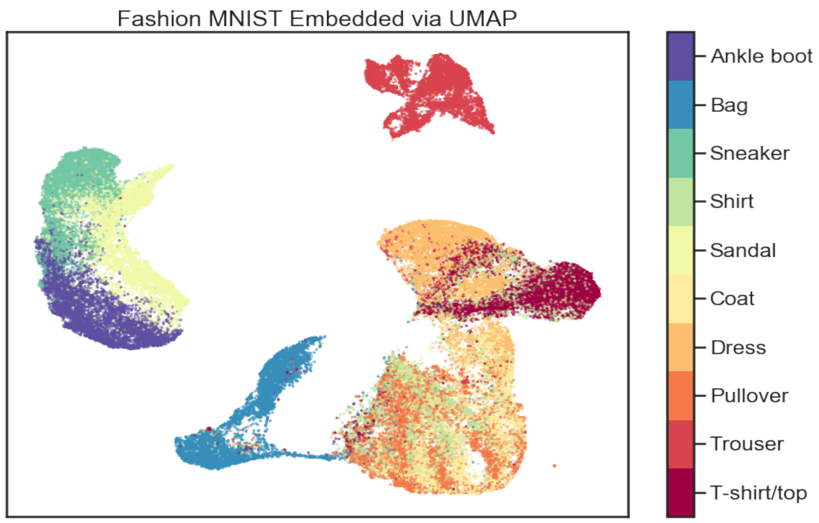

Рис. 1: Алгоритм UMAP

$p(x)$ — некоторая функция, определяющая вероятность того, что $X = 1$. Запишем формулу D_{KL} для X:

$$D_{KL}(P\,||\,\tilde{P}) = -p(x)\log\tilde{p}(x) - (1-p(x))\log(1-\tilde{p}(x)) + p(x)\log p(x) + (1-p(x))\log(1-p(x)) =$$

$$= p(x)\log\frac{p(x)}{\tilde{p}(x)} + (1-p(x))\log\frac{1-p(x)}{1-\tilde{p}(x)}$$

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

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

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 $S(P||\tilde{P})$ 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 $D_{KL} = CE(P||\tilde{P}) - H(p)$, то видно, что энтропия не зависит от $\tilde{p}(x)$, соответственно, является константой при минимизации. Тогда задача преобразуется в оптимизацию лишь суммы кросс-энтропий:

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

Поэтому UMAP считается методом, основанным на кросс-энтропии.

При отображении пространства 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 $t_1(m)$:

$$S(E_h||E_l)_1 = 0.7\log\frac{0.7}{0.6} + (1-0.7)\log\frac{1-0.7}{1-0.6} + 0.1\log\frac{0.1}{0.3} + (1-0.1)\log\frac{1-0.1}{1-0.3} +$$

$$+0.3\log\frac{0.3}{0.5} + (1-0.3)\log\frac{1-0.3}{1-0.5} \approx 0.22$$

Кросс-энтропия для $t_2(m)$:

$$S(E_h||E_l)_2 = 0.7\log\frac{0.7}{0.1} + (1-0.7)\log\frac{1-0.7}{1-0.1} + 0.1\log\frac{0.1}{0.9} + (1-0.1)\log\frac{1-0.1}{1-0.9} +$$

$$+0.3\log\frac{0.3}{0.4} + (1-0.3)\log\frac{1-0.3}{1-0.4} \approx 2.81$$

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. Воспользуемся формулой стандартной энтропии. Вероятность того, что бабушка находится в какой-либо одной конкретной деревне из трех равна $1/3$. Следовательно, энтропия местонахождения бабушки Елены в таком случае равна:

$$H = -\sum_{i=1}^{3}\frac{1}{3}\log\frac{1}{3}$$

Результатом подсчёта данного выражения будет число 1.58. Это означает, что в среднем (при большом количестве повторений эксперимента) нам понадобится 1.58 вопросов, чтобы верно назвать деревню. При округлении до ближайшего целочисленного даёт значение два. Оно и верно! Смотрите, можно узнать, в какой деревне живёт бабушка на 100 % всего за два вопроса. Например, 1) Название деревни - это гласная буква? (если да, то ответ уже найден; если нет следует задать еще один вопрос); 2) Название деревни начинается на букву Б? (в любом случае, далее вы уже сможете ответить на вопрос правильно с вероятностью, равной единице).
В целом можем сказать, что в среднем нам хватит двух вопросов, чтобы угадать деревню, а вот одного вопроса может хватить не всегда.

Ответ: 1.58

Задача 2. РЕШЕНИЕ И ОТВЕТ

Задача 3. Сюрприз! Вам предоставляется невероятная возможность самостоятельно проверить свои силы и решить следующие задачи. Варианты решения мы просим присылать в issues на открытый репозиторий на сайте github.com[4]. Первое правильное решение поощряется специальным ценным призом.

Желаем удачи и с нетерпением ждем Ваших решений!

Мы предоставляем ответы к задачам 3-5.

Ответ:

Задача 4. Ответ:

Задача 5. Ответ:

Задача 6. РЕШЕНИЕ и ОТВЕТ

Задача 7. РЕШЕНИЕ и ОТВЕТ

Задача 8. Сюрприз! Вам предоставляется невероятная возможность самостоятельно проверить свои силы и решить следующие задачи. Варианты решения мы просим присылать в issues на открытый репозиторий на сайте github.com. Первое правильное решение поощряется специальным ценным призом.

Желаем удачи и с нетерпением ждем Ваших решений!

Мы предоставляем ответы к задачам 8-10.

Ответ:

Задача 9. Ответ:

[4]Ссылка на репозиторий: `https://github.com/oomaystrenko/entropy`