

# Прикладная статистика в машинном обучении Семинары

Борис Демешев, Антон Золотарев, Дарья Демидова, Максим Каледин

Осень 2022

# Оглавление

<b>1</b>	<b>Энтропия и KL-дивергенция</b>	<b>3</b>
1.1	Разбираемся с энтропией . . . . .	4
1.2	Энтропия посложнее . . . . .	4
1.3	Сравнение энтропий двух распределений . . . . .	4
1.4	KL неотрицательна . . . . .	5
1.5	Энтропия ограничена . . . . .	5
1.6	Данетки . . . . .	5
1.7	У нормального распределения максимальная энтропия в классе распределений с фиксированным матожиданием и дисперсией . . . . .	6
1.8	Кросс-энтропия дискретного и непрерывного распределений . . . . .	6

# Энтропия и KL-дивергенция

$$H(X) = - \sum_i p_i \log_2(p_i)$$

$$H(X) = - \int_a^b f(x) \log_2(f(x)) dx$$

$$H(X, Y) = - \sum_{x, y} p(x, y) \log_2 p(x, y)$$

$$CE(X||Y) = - \sum p(x) \log_2 q(x)$$

$$H(X|Y) = H(X, Y) - H(Y)$$

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

$$D_{KL}(X||Y) = CE(X||Y) - H(X) = \sum_{x \in X} p(x) \log_2 \frac{p(x)}{p(y)}$$

## 1.1 Разбираемся с энтропией

### Условие задачи

Пусть  $X$  – случайная величина. Найдите  $H(X)$ , если

[а]  $X$  равновероятно принимает значения 1, 5, 7.

[б]  $X$  равномерно распределена на  $[0; a]$ .

[в]  $X \sim \mathcal{N}(\mu, \sigma^2)$ .

### Решение

будет!

## 1.2 Энтропия посложнее

### Условие задачи

Пусть  $S$  – радемахеровская случайная величина (принимает значения в  $\pm 1$  с вероятностями 0.5), независимая от  $X \sim \text{Binomial}(4, 0.5)$ . Найдите энтропию  $H(S(X + 1))$ .

### Решение

будет!

## 1.3 Сравнение энтропий двух распределений

### Условие задачи

В двух островных государствах Аббавг и Габбавг используют азбуку Морзе, каждая для своего национального языка. Каждый язык использует буквы А,Б,В,Г и они возникают с вероятностями как ниже.

Буква, $x$	А	Б	В	Г
$p_i$	0.1	0.15	0.45	0.3
Буква, $y$	А	Б	В	Г
$q_i$	0.2	0.4	0.2	0.2

Таблица 1.1: Частоты возникновения букв в аббавгском  $p_i$  и в габбавгском  $q_i$  языках

Научный прогресс в государствах позволил для каждого из языков построить оптимальный код для азбуки Морзе. В каком языке в среднем кодовые слова будут длиннее? Насколько отличаются средние длины кодовых слов в азбуках Морзе Аббавгского и Габбавгского?

### Решение

будет!

## 1.4 KL неотрицательна

### Условие задачи

(Неравенство Гиббса) Пусть даны два распределения вероятностей на конечном или счётном множестве  $I$ ,  $(p_i)$  и  $(p'_i)$ . Тогда для любого  $b > 1$  верно

$$\sum_{i \in I} p_i \log_b \frac{p'_i}{p_i} \leq 0.$$

### Решение

будет!

## 1.5 Энтропия ограничена

### Условие задачи

Пусть дискретные случайные величины  $X$  и  $Y$  принимают  $m$  значений. Докажите, что

1. KL-дивергенция  $D_{KL}(X \| Y)$  неотрицательна (используйте неравенство Гиббса) и что
2. для энтропии верно неравенство

$$0 \leq H(X) \leq \log_2 m.$$

### Решение

будет!

## 1.6 Данетки

### Условие задачи

Шерлок Холмс и доктор Ватсон играют в данетки; Холмс утверждает, что сможет отгадать в тему первой полосы газеты, которую лежит верхней в стопке. Оба примерно представляют список возможных тем: (1)постройка нового завода, (2)прокладка новой железной дороги, (3)происшествия в порту, (4)политический кризис, (5)новый локальный конфликт на Ближнем Востоке. Ватсон уверен, что с большой вероятностью 0.8 на первой полосе случится новый конфликт на Ближнем Востоке, а остальные события едва ли происходят летом, поэтому он верит в них на 0.05. У Холмса же есть свой бесценный опыт и экспертиза, который он бережёт и обогащает, он считает, что в текущем году примерно так:

- (1) и (2) освещают на первой полосе в 3 случаях из 10 каждый;
- (3) происходит редко: 1 случая из 20;
- (4) бывает, может, в 1 случае из 10;
- (5) в 1 из 4 случаев: ну что ж, такой регион.

Удивительно, но Холмс оказывается в таких вопросах нередко оказывается невероятно точен.

Сколько вопросов нужно задать Холмсу, чтобы угадать тему, если опыт его не подводит?

Сколько вопросов в среднем должен задать Ватсон, чтобы угадать тему при условии, что опыт Холмса не подводит? За сколько вопросов справится Холмс, если его опыт не релевантен и на самом деле прав Ватсон?

### **Решение**

будет!

**1.7 У нормального распределения максимальная энтропия в классе распределений с фиксированным матожиданием и дисперсией**

**1.8 Кросс-энтропия дискретного и непрерывного распределений**