

# Олимпиада ДЛС. Отборочный этап для студентов+

[Положение участников](#) [Задачи](#) [Ответы](#)

13 июн 2024, 18:46:41

старт: 11 мар 2024, 14:33:46

Объявления жюри

Завершить

...

## A. Хитрый интеграл (5)

Какое наибольшее значение может принимать функционал

$$I(p) = \int_0^{\infty} \exp\left(\frac{(x-4)^2 - (x+4)^2}{4}\right) \ln p(x) dx$$

при условии  $p(x) > 0, \int_0^{\infty} p(x) dx = 1$ ?

### Формат вывода

Введите ответ в виде десятичной дроби, округлённой до трёх знаков после запятой. В качестве разделителя используйте точку.

[Отправить](#)

осталось 3 попытки

[A. Хитрый интеграл \(5\)](#)[B. Неоднозначный РСА \(5\)](#)[C. Метод ближайших соседей \(5\)](#)[D. Градиентный спуск \(5\)](#)[E. Многоклассовая логистическая регрессия \(8\)](#)[F. Обучение решающего дерева \(8\)](#)[G. Свёрточные нейросети \(8\)](#)[H. Мандаринки \(8\)](#)[I. Сколько стоит дом \(16\)](#)[J. Судьба стартапа \(16\)](#)[K. Посчитать животных \(16\)](#)

Постылок нет

[Следующая](#)

# Олимпиада ДЛС. Отборочный этап для студентов+

[Положение участников](#) [Задачи](#) [Ответы](#)

## В. Неоднозначный РСА

Известно, что алгоритм [PCA](#) не всегда даёт однозначный ответ, потому что максимальный разброс может достигаться в проекции на несколько различных направляющих векторов.

Дан набор данных из 8 точек в трёхмерном пространстве, где точки являются вершинами 3-мерного куба. Для этого набора данных обучили РСА с одной главной компонентой (`n_components=1` в библиотеке `sklearn`). Какое количество различных ответов могло получиться? Иными словами, найдите количество направлений, проекции точек 3-мерного куба на которые имеют максимальный разброс. Направления, отличающиеся только знаком направляющего вектора, считаются различными.

- 3
- 4
- 6
- 8
- 12
- 16
- бесконечно

[Отправить](#)

? осталось 2 попытки

[Предыдущая](#)

13 июн 2024, 18:47:43

старт: 11 мар 2024, 14:33:46

[Объявления жюри](#)[Завершить](#)

...

A. Хитрый интеграл (5)

B. Неоднозначный РСА (5)

C. Метод ближайших соседей (5)

D. Градиентный спуск (5)

E. Многоклассовая логистическая регрессия (8)

F. Обучение решающего дерева (8)

G. Свёрточные нейросети (8)

H. Мандаринки (8)

I. Сколько стоит дом (16)

J. Судьба стартапа (16)

K. Посчитать животных (16)

# Олимпиада ДЛС. Отборочный этап для студентов+

[Положение участников](#) [Задачи](#) [Ответы](#)

13 июн 2024, 18:48:49

старт: 11 мар 2024, 14:33:46

Объявления жюри

Завершить

...

## С. Метод ближайших соседей

Рассмотрим  $\ell$  точек, распределенных равномерно по объему  $n$ -мерного единичного шара с центром в нуле. Предположим, что мы хотим применить метод одного ближайшего соседа для точки  $O$  — начала координат. Зададимся вопросом, на каком расстоянии будет расположен ближайший объект к точке  $O$ . Для ответа на этот вопрос выведите выражение для медианы расстояния от  $O$  до ближайшего объекта.

Дайте ответ для  $\ell = 500$  и  $n = 10$

### Формат вывода

В ответ запишите десятичную дробь, округлив её до 3 знаков после запятой. Используйте точку в качестве разделителя.

[Отправить](#) [осталось 3 попытки](#)[Предыдущая](#)[Следующая](#) [Посылок нет](#)[A. Хитрый интеграл \(5\)](#)[B. Неоднозначный РСА \(5\)](#)[C. Метод ближайших соседей \(5\)](#)[D. Градиентный спуск \(5\)](#)[E. Многоклассовая логистическая регрессия \(8\)](#)[F. Обучение решающего дерева \(8\)](#)[G. Свёрточные нейросети \(8\)](#)[H. Мандаринки \(8\)](#)[I. Сколько стоит дом \(16\)](#)[J. Судьба стартапа \(16\)](#)[K. Посчитать животных \(16\)](#)

# Олимпиада ДЛС. Отборочный этап для студентов+

[Положение участников](#) [Задачи](#) [Ответы](#)

13 июня 2024, 18:49:18

старт: 11 мар 2024, 14:33:46

Объявления жюри

Завершить

...

## D. Градиентный спуск (5)

Пусть мы ищем минимум функции  $f(x) = x^4$  с помощью градиентного спуска. Стартовая точка  $x_0 = 4$ . Используется обычный градиентный спуск, шаг которого выражается формулой

$$x_{i+1} = x_i - \alpha \cdot \frac{df}{dx}(x_i).$$

При каком минимальном значении  $\alpha > 0$  алгоритм не сойдется?

Иными словами, при каком минимальном значении  $\alpha > 0$  будет существовать такое  $\delta > 0$ , что алгоритм градиентного спуска, стартуя из точки  $x_0$ , никогда не придет в интервал  $(-\delta, \delta)$ ?

### Формат вывода

Дайте ответ в виде несократимой дроби  $p / q$ .

[Отправить](#)

осталось 2 попытки

[Предыдущая](#)[Следующая](#)

- A. Хитрый интеграл (5)
- B. Неоднозначный РСА (5)
- C. Метод ближайших соседей (5)
- D. Градиентный спуск (5)
- E. Многоклассовая логистическая регрессия (8)
- F. Обучение решающего дерева (8)
- G. Свёрточные нейросети (8)
- H. Мандаринки (8)
- I. Сколько стоит дом (16)
- J. Судьба стартапа (16)
- K. Посчитать животных (16)

# Олимпиада ДЛС. Отборочный этап для студентов+

13 июня 2024, 18:49:48

старт: 11 мар 2024, 14:33:46

Объявления жюри

Завершить

...

Положение участников Задачи Ответы

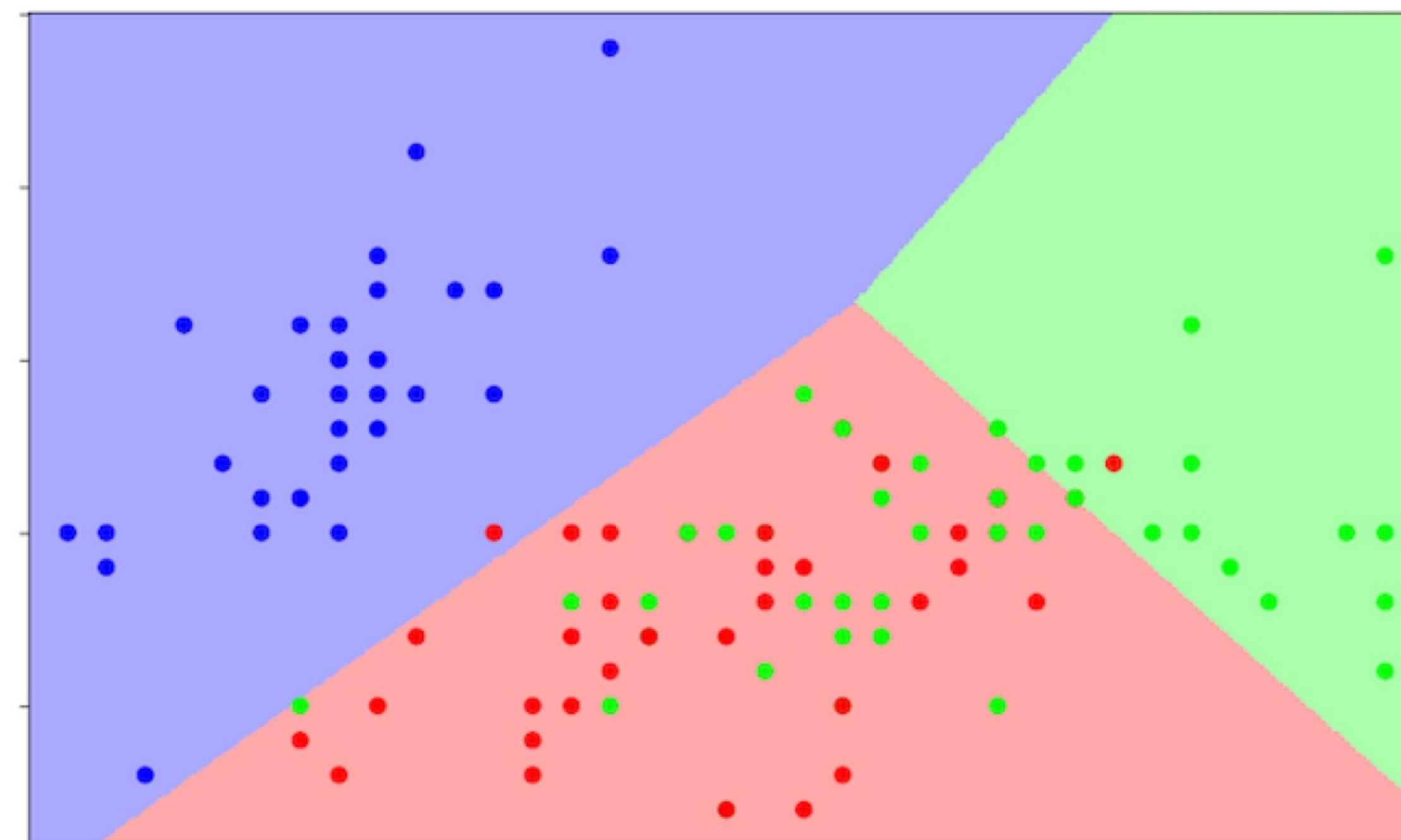
## E. Многоклассовая логистическая регрессия (8)

Для задачи классификации на  $k$  классов  $\{1, \dots, k\}$  обучили алгоритм многоклассовой логистической регрессии, который устроен следующим образом. Если объекты в датасете имеют  $n$  признаков, то обучаемыми параметрами алгоритма являются матрица  $W$  размера  $k \times n$  и вектор  $b$  размера  $k$ . Вектор вероятностей принадлежности классам  $1, \dots, k$  вычисляется по формуле

$$p = \text{softmax}(Wx + b).$$

На картинке изображено, как классификатор разбивает пространство признаков на три части, соответствующие классам. Здесь  $n = 2$ ,  $k = 3$ ,  $W = \begin{pmatrix} -3 & 3 \\ 1 & -2 \\ 2 & -1 \end{pmatrix}$ ,  $b = \begin{pmatrix} 8 \\ 1 \\ -9 \end{pmatrix}$ .

Пусть  $A$  --- точка пересечения всех трёх линий разделения классов. Найдите координаты точки  $A$ .



A. Хитрый интеграл (5)

B. Неоднозначный РСА (5)

C. Метод ближайших соседей (5)

D. Градиентный спуск (5)

E. Многоклассовая логистическая регрессия (8)

F. Обучение решающего дерева (8)

G. Свёрточные нейросети (8)

H. Мандаринки (8)

10.5 I. Сколько стоит дом (16)

4.41 J. Судьба стартапа (16)

14.76 K. Посчитать животных (16)

### Формат вывода

Ответ дайте в виде двух несократимых дробей вида  $p/q$  через пробел. Например: "1/2 1/3".

Отправить

?

осталась 1 попытка

Предыдущая

Следующая

# Олимпиада ДЛС. Отборочный этап для студентов+

Положение участников Задачи Ответы

13 июня 2024, 18:50:22

Объявления жюри

Завершить

старт: 11 мар 2024, 14:33:46

...

## F. Обучение решающего дерева (8)

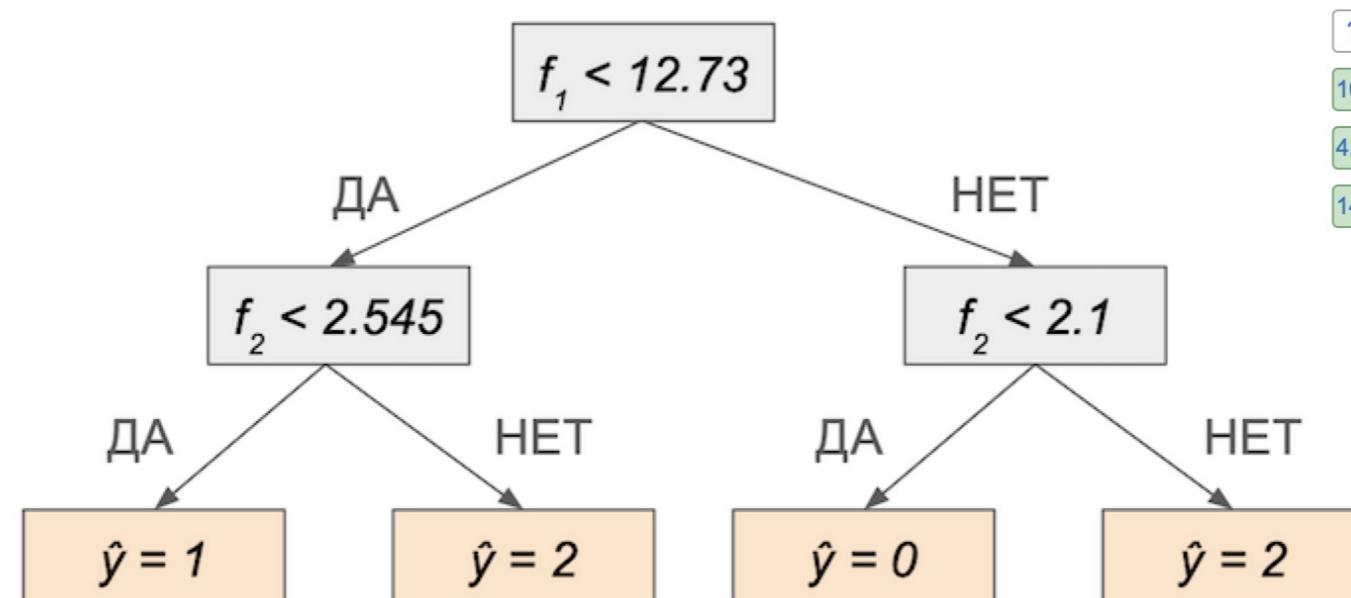
Пусть мы решаем задачу классификации на 3 класса, и в нашем обучающем датасете [train.csv](#) 18 элементов. Мы хотим построить для решения этой задачи решающее дерево с максимальной глубиной 2.

В качестве критерия информативности (impurity function), выберем следующий критерий:

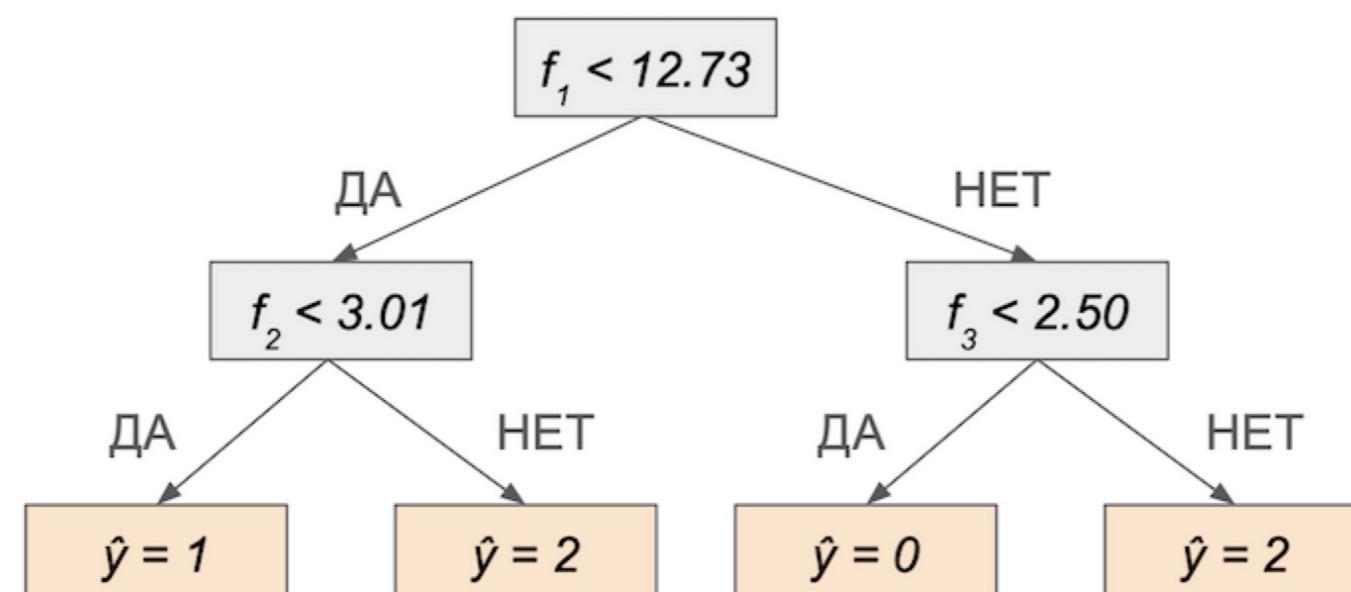
$$GI = \sum_{i=1}^3 p_i(1 - p_i), \text{ где } p_i \text{ — доля элементов, принадлежащих классу } i$$

Какое из представленных ниже решающих деревьев может получиться при таких параметрах?

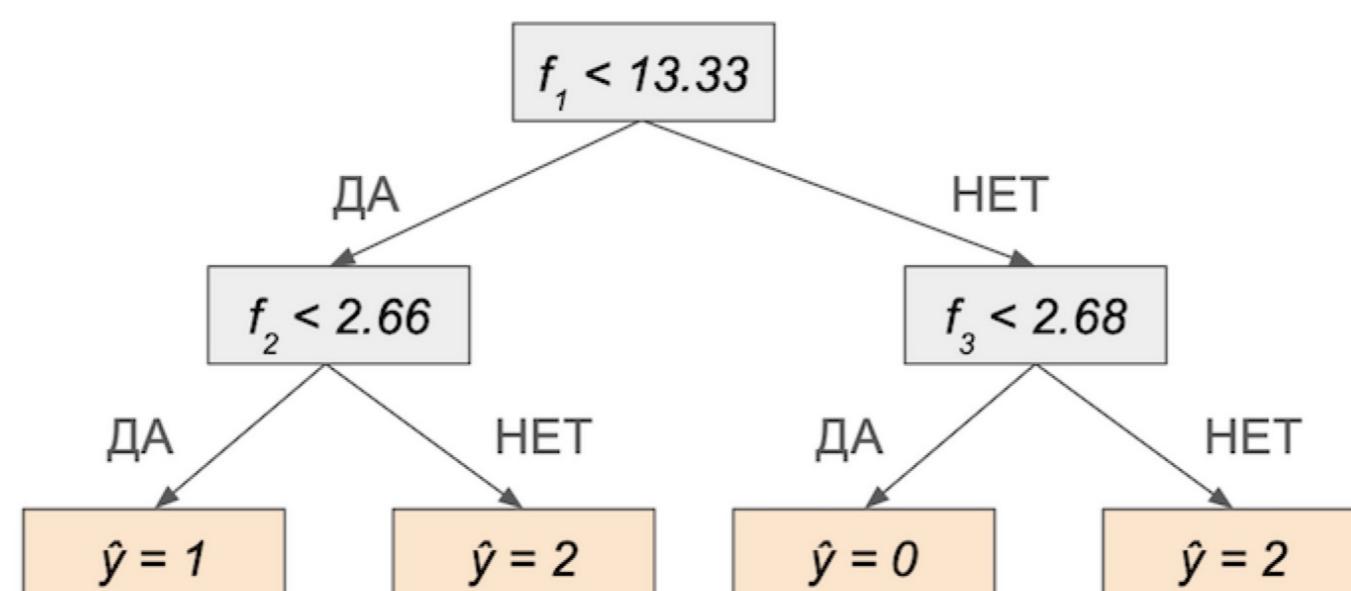
Дерево №1



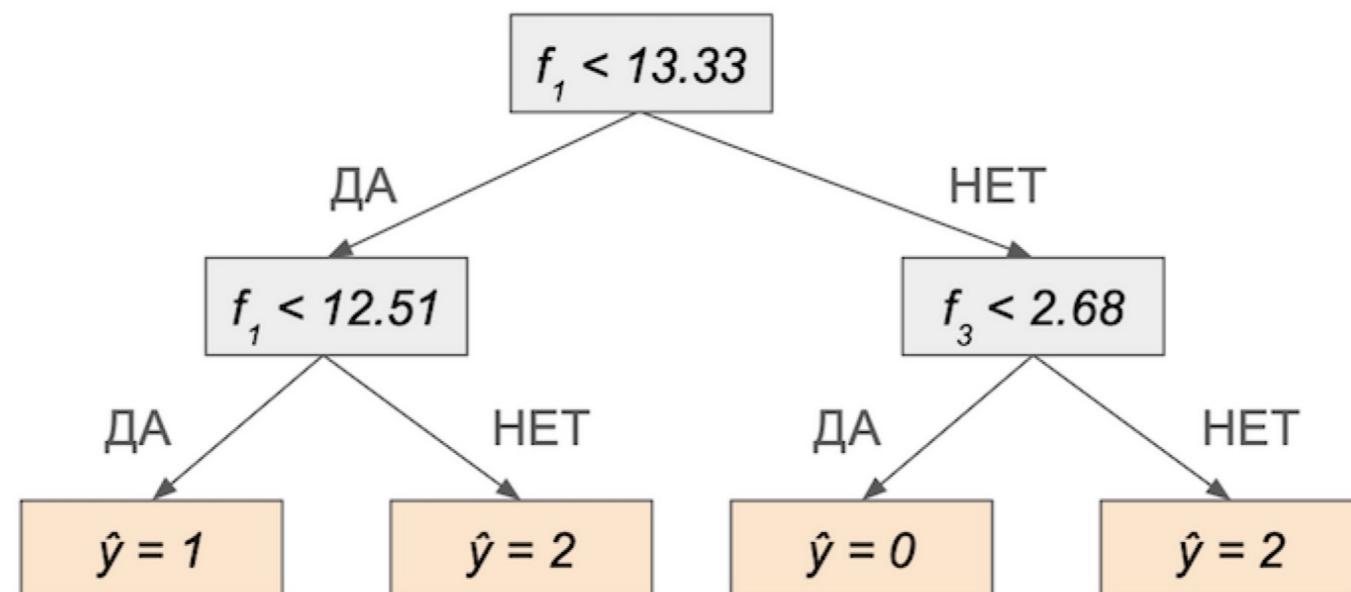
Дерево №2



Дерево №3



Дерево №4



### Примечания

[Описание алгоритма построения решающего дерева](#)

- 1
- 2
- 3
- 4



осталось 2 попытки

# Олимпиада ДЛС. Отборочный этап для студентов+

13 июня 2024, 18:50:47

старт: 11 мар 2024, 14:33:46

Объявления жюри

Завершить

...

Положение участников Задачи Ответы

## G. Свёрточные нейросети (8)

Таня решает задачу классификации изображений размера  $3 \times 3$  на два класса. В обучающей выборке три изображения:

1	2	3
3	4	1
2	1	3

класс 1

1	1	2
2	6	1
0	1	2

класс 2

2	2	1
2	4	3
1	2	4

класс 1

Таня придумала следующий алгоритм классификации:

- Сворачиваем изображение фильтром размера  $2 \times 2$ . Получаем карту активации размера  $2 \times 2$ .
- Применяем к карте активации MaxPooling с ядром 2. Получаем одно число  $C$ .
- Выбираем число  $thr$ . Если  $C > thr$ , то изображение считается принадлежащим классу 2, если  $C \leq thr$ , то изображение считается принадлежащим классу 1.

Таня хочет подобрать такой фильтр, чтобы придуманный ею алгоритм идеально делил элементы выборки на два класса. Какой из представленных ниже фильтров подойдет Тане? Иными словами, при использовании какого из фильтров, представленных ниже, в алгоритме Тани получится подобрать значение  $thr$ , чтобы алгоритм идеально классифицировал три изображения?

1	1
1	1

1	0
0	1

0	1
0	1

0	1
1	0

фильтр 1

фильтр 2

фильтр 3

фильтр 4

- 1
- 2
- 3
- 4

**Отправить**

?

осталось 2 попытки

Предыдущая

Следующая

A. Хитрый интеграл (5)

B. Неоднозначный РСА (5)

C. Метод ближайших соседей (5)

D. Градиентный спуск (5)

E. Многоклассовая логистическая регрессия (8)

F. Обучение решающего дерева (8)

G. Свёрточные нейросети (8)

H. Мандаринки (8)

I. Сколько стоит дом (16)

J. Судьба стартапа (16)

K. Посчитать животных (16)

# Олимпиада ДЛС. Отборочный этап для студентов+

13 июня 2024, 18:51:09

старт: 11 марта 2024, 14:33:46

Объявления жюри

Завершить

...

Положение участников Задачи Ответы

## Н. Мандаринки (8)

Тася и ее друг Ваня очень любят мандарины, но Тася предпочитает абхазские, а Ваня — марокканские. Многочисленные эксперименты с мандаринами позволили ребятам вывести четыре основных показателя внешнего вида мандарина:

- толщина мандарина  $x_1$  (численный признак);
- диаметр наибольшего сечения мандарина  $x_2$  (численный признак);
- насыщенность цвета  $x_3$  (численный признак);
- наличие/отсутствие листика  $x_4$  (бинарный признак, принимает значения 0/1).

Численные признаки приведены в особых единицах, поэтому могут принимать как положительные, так и отрицательные значения. Более того, ребята установили, что распределение вектора параметров  $(x_1, x_2, x_3, x_4)$  следующее: для обоих сортов вектор  $(x_1, x_2, x_3)$  имеет трехмерное нормальное распределение, причём

- для абхазских мандаринов — с вектором средних  $a_1 = (0, 0, 0)$  и матрицей ковариаций

$$K_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix};$$

- для марокканских — с вектором средних  $a_2 = (6, 6, 6)$  и матрицей ковариаций

$$K_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Наконец, признак  $x_4$  не зависит от совокупности признаков  $(x_1, x_2, x_3)$  для обоих сортов, причём для абхазского имеет распределение Бетн ( $p_1$ ), а для марокканского — Бетн ( $p_2$ ) (здесь  $p_i$  — вероятность наличия листика у мандаринки  $i$ -ого класса,  $i = 1, 2$ ),  $p_1 = 1 / 2$ ,  $p_2 = 1 / 3$ .

Тася купила мандарины двух видов, оставила сумки на столе и пошла отдохнуть, но, пока она спала, кошка Мурка взобралась на стол, вывалила мандарины из обеих сумок и перемешала их. Помогите ребятам построить дискриминативную функцию методом максимального правдоподобия!

### Формат ввода

Вам дан файл [tangerines.csv](#), в котором даны параметры 10 мандаринов.

### Формат вывода

Вам необходимо сдать последовательность из 10 чисел: 1 или 2, где 1 означает, что для данного мандарина выше правдоподобие того, что он абхазский, а 2 — марокканский.

A. Хитрый интеграл (5)

B. Неоднозначный РСА (5)

C. Метод ближайших соседей (5)

D. Градиентный спуск (5)

E. Многоклассовая логистическая регрессия (8)

F. Обучение решающего дерева (8)

G. Свёрточные нейросети (8)

H. Мандаринки (8)

I. Сколько стоит дом (16)

J. Судьба стартапа (16)

K. Посчитать животных (16)

**Отправить**

?

осталось 2 попытки

Предыдущая

Следующая

# Олимпиада ДЛС. Отборочный этап для студентов+

13 июня 2024, 18:52:07  
старт: 11 мар 2024, 14:33:46  
...

[Объявления жюри](#)[Завершить](#)[Положение участников](#) [Задачи](#) [Ответы](#)

## I. СКОЛЬКО СТОИТ ДОМ

Петин друг Вася — риэлтор, занимается продажами квартир. Для каждой квартиры он вручную оценивает ее стоимость по ее параметрам. Петя занимается машинным обучением, и решил помочь Васе в его работе. Петя хочет обучить модель машинного обучения, которая по данным о квартире предсказывала бы ее стоимость. Помогите Пете это сделать.

### Формат ввода

#### [Ссылка на данные](#)

Вам даны тренировочная и тестовая выборки данных с данными о недвижимости в Москве и Санкт-Петербурге. Тренировочная выборка содержит следующие столбы:

- id — уникальный идентификатор квартиры;
- author — автор объявления;
- author\_type — тип автора объявления (агент, застройщик и т.п.);
- location — город;
- deal\_type — тип продажи;
- accommodation\_type — тип объекта недвижимости;
- floor — этаж, на котором находится объект;
- floors\_count — сколько всего этажей в доме;
- rooms\_count — количество комнат;
- total\_meters — площадь в квадратных метрах;
- district — район;
- street — улица;
- underground — станция метро;
- residential\_complex — название жилого комплекса, в котором находится объект;
- price (целевая переменная) — цена в миллионах рублей.

[A. Хитрый интеграл \(5\)](#)[B. Неоднозначный РСА \(5\)](#)[C. Метод ближайших соседей \(5\)](#)[D. Градиентный спуск \(5\)](#)[E. Многоклассовая логистическая регрессия \(8\)](#)[F. Обучение решающего дерева \(8\)](#)[G. Свёрточные нейросети \(8\)](#)[H. Мандаринки \(8\)](#)[10.5 I. Сколько стоит дом \(16\)](#)[4.41 J. Судьба стартапа \(16\)](#)[14.76 K. Посчитать животных \(16\)](#)

### Формат вывода

Тестовая выборка содержит все те же столбцы, кроме целевой переменной. Ваша задача — обучить модель машинного обучения на тренировочных данных и предсказать величину *price* для тестовых данных. В качестве решения нужно отправить ответы модели на тестовой выборке. Пример файла ответа можно найти по [ссылке](#).

### Примечания

Ваше решение будет оцениваться как задача регрессии по метрике *MAE* (Mean Absolute Error).

Количество баллов за посылку равно

$$\frac{60 - MAE}{3.75}$$

где *MAE* — значение метрики МАЕ для вашего ответа.

[Набрать здесь](#) [Отправить файл](#)

1

[Отправить](#)

10.5

осталось 18 попыток

[Предыдущая](#)[Следующая](#)

# Олимпиада ДЛС. Отборочный этап для студентов+

13 июня 2024, 18:52:41

старт: 11 марта 2024, 14:33:46

Объявления жюри

Завершить

...

Положение участников Задачи Ответы

## J. Судьба стартапа (16)

Андрей хочет запустить свой стартап. Но Андрей знает, что многие стартапы не доходят до успеха и закрываются. Чтобы не прогореть, Андрей решил выяснить, что отличает хороший стартап от плохого, с помощью машинного обучения. Для этого Андрей собрал информацию о стартапах и решил обучить алгоритм машинного обучения определять, выживет стартап или нет. Помогите Андрею это сделать!

### Формат ввода

#### Ссылка на данные

Вам даны тренировочная и тестовая выборки данных с данными о стартапах. Тренировочная выборка содержит следующие столбцы:

- index — уникальный id стартапа;
- name — название стартапа;
- overview — описание стартапа;
- category\_code — категория, в которой работает стартап;
- tag\_list — список тэгов для стартапа;
- country\_code — страна, в которой зарегистрирован стартап;
- num\_prev\_rounds — количество раундов, который к настоящему времени поднял стартап. Раунд — это получение денег от инвесторов;
- has\_raised\_amount — получал ли стартап до настоящего момента деньги от инвесторов;
- ln\_raised\_amount — логарифм суммы денег, которые стартап получил от инвесторов к настоящему моменту;
- participants — количество инвесторов, инвестировавших деньги в стартап;
- previous\_any\_founder\_experience — был ли у основателя стартапа опыт в работе/бизнесе до основания стартапа;
- ipo\_prob — вычисленное Андреем значение вероятности, что в будущем стартап попробует стать публичной компанией;
- ma\_prob — вычисленное Андреем значение, что в будущем стартап будет приобретен другой компанией;
- has\_next\_round (целевая переменная) — выживет ли стартап.

A. Хитрый интеграл (5)

B. Неоднозначный РСА (5)

C. Метод ближайших соседей (5)

D. Градиентный спуск (5)

E. Многоклассовая логистическая регрессия (8)

F. Обучение решающего дерева (8)

G. Свёрточные нейросети (8)

H. Мандаринки (8)

I. Сколько стоит дом (16)

J. Судьба стартапа (16)

K. Посчитать животных (16)

### Формат вывода

Тестовая выборка содержит все те же столицы, кроме целевой переменной. Ваша задача — обучить модель машинного обучения на тренировочных данных и предсказать величину *has\_next\_round* для тестовых данных. В качестве решения нужно отправить ответы модели на тестовой выборке. Пример файла ответа можно найти по [ссылке](#).

### Примечания

Ваше решение будет оцениваться как задача классификации по метрике *Accuracy*.

Количество баллов за посылку равно

$$32 * (Acc - 0.5)$$

где *Acc* — значение метрики Accuracy для вашего ответа.

Набрать здесь Отправить файл

1	
---	--

Отправить

4.41

осталось 18 попыток

Предыдущая

Следующая

# Олимпиада ДЛС. Отборочный этап для студентов+

⌚ 13 июня 2024, 18:53:06  
старт: 11 марта 2024, 14:33:46

[Объявления жюри](#)[Завершить](#)

...

[Положение участников](#) [Задачи](#) [Ответы](#)

## К. Посчитать животных (16)

Лиза очень любит коров и других животных. Она просит вас посчитать, сколько животных на изображениях, которые ей прислали из африканской саванны.

### Формат ввода

#### [Ссылка на данные](#)

Данные представляют из себя изображения с путями [train,test]/pictures/[train,test]\_{n}.png. Размер каждого изображения --- 100 × 100. В файле train/answers.csv указаны правильные ответы: количество животных на каждом изображении.

### Формат вывода

Вам необходимо послать в соревнование файл submission.csv,  $k + 1$ -ая строка которого будет содержать предсказанное количество животных на изображении k.png. Обратите внимание, что предсказания не обязаны быть целыми числами. Нулевая строка должна содержать название колонки. Файл sample\_submission.csv представляет из себя пример посылки. Файл test/image\_names.csv содержит список тестовых изображений в том порядке, в котором необходимо перечислить ответы для них в submission.csv

### Примечания

Метрика оценки качества --- Mean Absolute Error, которая вычисляется по формуле

$$\frac{1}{\ell} \sum_{k=1}^{\ell} |y_{true} - y_{pred}|.$$

Оценка ставится по формуле

$$16 \cdot \frac{\text{BaselineScore} - \text{score}}{\text{BaselineScore}},$$

где score --- ваш результат, а BaselineScore = 1.

- [A. Хитрый интеграл \(5\)](#)
- [B. Неоднозначный РСА \(5\)](#)
- [C. Метод ближайших соседей \(5\)](#)
- [D. Градиентный спуск \(5\)](#)
- [E. Многоклассовая логистическая регрессия \(8\)](#)
- [F. Обучение решающего дерева \(8\)](#)
- [G. Свёрточные нейросети \(8\)](#)
- [H. Мандаринки \(8\)](#)
- [I. Сколько стоит дом \(16\)](#)
- [J. Судьба стартапа \(16\)](#)
- [K. Посчитать животных \(16\)](#)

[Набрать здесь](#) [Отправить файл](#)

1	
---	--

[Отправить](#)

14.76

осталось 19 попыток

[Предыдущая](#)