

- [Разное](#) > [Data Culture](#) > [Анализ данных](#)
- > [Подготовка к НЭ по Анализу данных.](#)
- [Продвинутый уровень](#)
- > Тренировочный вариант 1 UPDATE

Оставшееся время 2:44:07

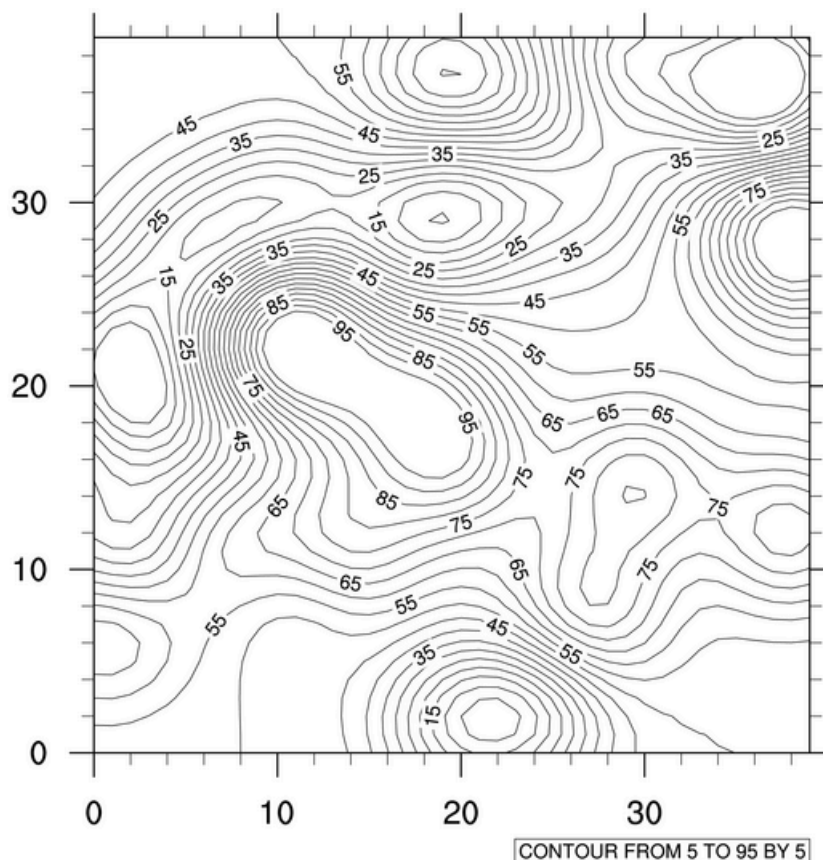
Вопрос 1Пока нет
ответа

Балл: 0,32

Иногда приходится иметь дело с трехмерными поверхностями, т.е. результатом табулирования функции от двух переменных $z = f(x, y)$. Сейчас, как правило, с восприятием таких графиков нет никаких проблем - есть много способов создать интерактивный трехмерный график, который можно вертеть и масштабировать как угодно. Но если мы говорим о статичной картинке, то с пониманием графика возникают некоторые трудности. Выход нашли ещё до появления компьютеров, при издании карт местности, который заключается в том, что строят не трехмерный график, а его отображение на плоскости. Это отображение получают по следующим правилам: берут заданное количество плоскостей, параллельных плоскости xy , и выделяют место пересечения поверхности с каждой из этих плоскостей - контуры, затем проецируют данные контуры на плоскость xy и всё — график готов.

Каждому контуру соответствует определенное значение величины z , его подписывают рядом с контуром.

Перед нами стоит задача минимизировать некоторую функцию потерь, contour plot которой изображен ниже.



Выберите все верные утверждения:

- ☐ В точке $(x,y)=(16,20)$ значение функции близко к своему минимальному значению.
- ☒ Функция имеет несколько локальных экстремумов
- ☒ На картинке есть как области локальных минимумов, так и области локальных максимумов

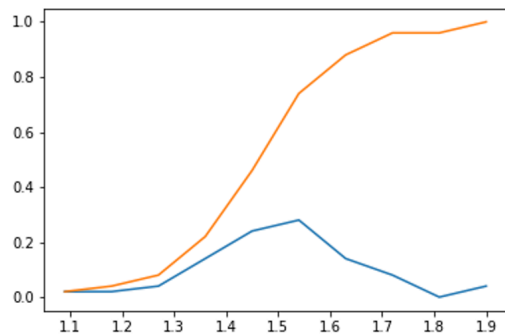
- ☐ По графику можно точно определить минимальное значение функции, оно равно 15

Вопрос 2

Пока нет ответа

Балл: 0,32

На картинке нарисованы графики плотности распределения (pdf) и функции распределения (cdf) некоторой случайной величины.



Выберите верный вариант ответа:

- ☐ Оранжевый график - это плотность, синий - функция распределения
- ☐ Мода случайной величины равна примерно 1.8
- ☐ Оранжевый график не может быть плотностью
- ☒ Оранжевый график - это функция распределения, синий - плотность

Очистить мой выбор

Вопрос 3

Пока нет ответа

Балл: 0,32

Аналитик Вася с утра просыпается с головной болью и попадает по нужной клавише на клавиатуре с вероятностью 0.7. Найдите математическое ожидание количества правильно нажатых клавиш в предложении из 40 символов.

Ответ:

Вопрос 4

Пока нет
ответа

Балл: 0,32

Инженер данных Сергей построил пайплайн для сбора и хранения данных. Сергей утверждает, что в его алгоритме данные теряются с вероятностью 2%. Затем пришел инженер данных Виталий и, проверяя гипотезу Сергея, заявил, что такая маленькая потеря данных невозможна. Сергей обиделся и протестировал свой алгоритм многократными проверками, при этом каждый раз только 2% данных терялось.

Выберите верное утверждение:

- ☐ Виталий совершил ошибку второго рода
- ☐ Виталий не совершил ошибку, ошибку совершил Сергей - но какого рода, неизвестно
- ☒ Виталий совершил ошибку первого рода

[Очистить мой выбор](#)

Вопрос 5

Пока нет
ответа

Балл: 0,32

Предположим, вы проводите исследование благосостояния граждан по странам. Имеющиеся у вас данные включают в себя уровень дохода, среднегодовую температуру в месте проживания респондента, степень удовлетворенности жизнью, ВВП на душу населения, возраст респондента, его рост и другие характеристики для 100 тысяч граждан 100 государств. Хотим построить модель, которая будет по данным для человека предсказывать уровень благосостояния - любое число на отрезке от 1 до 100 (где 1 - низкий уровень, а 100 - высокий уровень). Ответ может быть любым числом из отрезка $[1; 100]$, в том числе и нецелым.

Выберите все верные утверждения (их два):

- ☒ Будем работать над решением задачи регрессии
- ☐ Имеем дело с задачей классификации
- ☒ Объектом является гражданин страны
- ☐ Объектом в этой задаче является страна

Вопрос 6

Пока нет
ответа

Балл: 0,32

Цель банка - найти мошенников при совершении банковских операций (мошеннических транзакций на несколько порядков меньше, чем не мошеннических). Пусть при обучении классификации для выявления мошеннических транзакций в банке ассигасу на тренировочных данных получилось равным 0.9, а на тестовых - 0.85. О чем может говорить эта ситуация?

Для решения задачи использовали логистическую регрессию с Lasso-регуляризацией.

Выберите наиболее подходящий вариант ответа.

- ☐ Модель обучилась как надо (нет ни недообучения, ни переобучения)
- ☐ Модель переобучилась, так как качество на тесте ниже, чем на трейне.
- ☐ Модель недообучилась, так как качество на трейне низкое
- ☒ Неизвестно, переобучилась или недообучилась модель, так как выбрана плохая метрика для данной задачи

Очистить мой выбор

Вопрос 7

Пока нет
ответа

Балл: 0,32

Дан текст: "Но не каждый хочет что-то исправлять :("

После некоторой обработки получилось:

```
['но', '', 'не', '', 'каждый', '', 'хотеть', '', 'что-то', '',  
'исправлять', ':(\n']
```

Выберите все шаги, которые были сделаны с исходным текстом:

- ☐ Стемминг (stemming)
- ☒ Лемматизация (lemmatization)
- ☒ Токенизация (tokenization)
- ☐ Векторизация (vectorization)

Вопрос 8

Пока нет
ответа

Балл: 0,32

Мы решаем задачу классификации - определения по МРТ-снимкам, болен пациент некоторой болезнью или нет. Мы не хотим часто ошибаться, то есть называть больных здоровыми, но и здоровых называть больными мы также не хотим. В обучающих данных 8543 пациента, из которых 4100 больных, а остальные здоровые.

Какие метрики можно использовать для измерения качества модели в этой задаче?

- ☐ MAPE
- ☐ MSLE (mean squared logarithmic error)
- ☒ Accuracy
- ☒ AUC-PR (площадь под Precision-Recall кривой)
- ☒ F1-score

Вопрос 9

Пока нет
ответа

Балл: 0,32

Выберите два корректных утверждения про градиентный спуск:

- ☐ На каждом шаге алгоритма считается градиент от одного, случайно выбранного элемента.
- ☐ Градиентный спуск применяется для нахождения максимума функции потерь
- ☒ Правильный подбор шага градиентного спуска позволяет уменьшить количество шагов, необходимых для поиска минимума
- ☒ Если сделать длину шага градиентного спуска недостаточно маленькой, то алгоритм может разойтись.

Вопрос 10Пока нет
ответа

Балл: 0,32

Напомним, что существует формула для разложения ошибки модели на шум, смещение и разброс:

$$Err = Bias^2 + Var + \delta^2$$

$Bias^2$ - смещение модели, Var - разброс, δ^2 - шум в данных.

На некоторых данных обучили линейную регрессию. Как изменятся компоненты $Bias^2$ и Var ошибки линейной регрессии, если добавить в данные для обучения модели полиномиальные признаки степени 2?

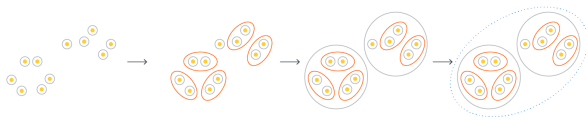
Среди перечисленных ниже два ответа верные.

- ☒ $Bias^2$ уменьшится
- ☐ $Bias^2$ увеличится
- ☐ Var не изменится
- ☐ Var уменьшится
- ☐ $Bias^2$ не изменится
- ☒ Var увеличится

Вопрос 11Пока нет
ответа

Балл: 0,32

На рисунке ниже изображен некоторый алгоритм кластеризации. А какой?



- ☐ Спектральная кластеризация
- ☒ Аггломеративная (иерархическая) кластеризация
- ☐ K-means
- ☐ DBSCAN

Очистить мой выбор


Предыдущий
элемент курса

Разбор Демоверсии
экзамена (видео)

Перейти на...

Следующий
элемент курса

Тренировочный
вариант 2 (из
реального экзамена)

 Служба
поддержки сайта

Вы зашли под именем
Царахова Милена
Викторовна (Выход)