Оставшееся время 2:58:30

Вопрос 1

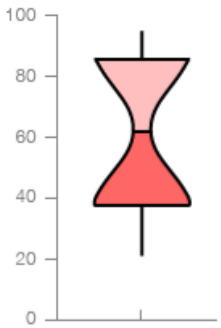
Пока нет ответа

Балл: 0,32

Vase plot - это тип визуализации данных, который сочетает в себе box plot (ящик с усами) и информацию о плотности распределения данных. Vase plot содержит в себе:

- Центральную горизонтальную линию - медиану данных
- Верхнюю и нижнюю горизонтали - 75% (Q3) и 25%(Q1)-квантили распределения данных
- Усы заканчиваются в наибольшем значении выборки, не превышающем Q3+1.5IQR, и наименьшем значении выборки, превышающем Q1-1.5IQR, где IQR=Q3-Q1
- Точки выше и ниже усов - выбросы
- Вертикальные стороны фигур, в отличие от классического box plot, это визуализация плотности распределения данных (её части на отрезке от 25% до 75% квантили).

На рисунке изображен vase plot, отображающий распределение веса в некоторой группе людей.



Выберите верные утверждения относительно отображенных данных:

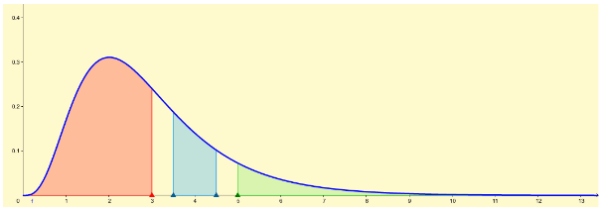
- [ ] В данных нет значений, меньших 30
- [x] Медиана набора данных приблизительно равна 60
- [x] В данных нет выбросов
- [ ] Данные имеют равномерное распределение
- [ ] Интерквартильный размах равен приблизительно 30

Вопрос 2

Пока нет ответа

Балл: 0,32

Время между клиентами (в минутах), посещающими магазин, имеет экспоненциальное распределение, показанное на рисунке ниже. По оси x отображено время в минутах.



По этому распределению посчитали три величины:

$P(X \leq 3) = 0.6, P(3.5 \leq X \leq 4.5) = 0.14, P(X \geq 5) = 0.1$

Выберите три верных утверждения.

- [x] Вероятность того, что следующий клиент придет на 7 или больше минут позже предыдущего, меньше 0.1
- [x] Вероятность того, что следующий клиент придет в течение первых трех минут после предыдущего, равна 0.6
- [x] Вероятность того, что следующий клиент придет на 4 или больше минут позже предыдущего, меньше 0.5
- [ ] Вероятность того, что следующий клиент придет в течение минуты после предыдущего, больше 0.6
- [ ] Каждый следующий клиент приходит не позднее, чем через 8 минут после предыдущего

Вопрос 3

Пока нет ответа

Балл: 0,32

Виктор пытается дозвониться в телепередачу на радио. Известно, что вероятность дозвониться равна 0.02 и не зависит от предыдущих попыток. Виктор дозвонился с 10й попытки. На сколько попыток раньше он дозвонился, чем в среднем дозваниваются желающие?

Ответ: 40

Вопрос 4

Пока нет ответа

Балл: 0,32

Имеется исследование о влиянии нового учебного метода на успеваемость студентов. Учебный метод внедрялся в группе студентов, исследователи ожидали, что он приведет к улучшению успеваемости.

Нулевая гипотеза: новый учебный метод не влияет на успеваемость студентов.

Пороговое значение статистической значимости (α) установлено на уровне 0.05. После анализа данных исследователи получили p-value = 0.08.

Какую ошибку исследователи могли допустить?

- ○ Обе ошибки
- ○ Ошибка первого рода
- ○ Ни одной ошибки
- ◉ Ошибка второго рода

Очистить мой выбор

Вопрос 5

Пока нет ответа

Балл: 0,32

Аналитики некоторой компании проводят исследование спроса клиентов на шариковые ручки.

Данные представлены в виде таблицы с полями "год покупки ручек", "возраст", "пол", "уровень образования", "средний доход", "город проживания клиентов".

Аналитики пытаются спрогнозировать количество шариковых ручек, которое каждый клиент купил за 2018, 2019, 2020, 2021, 2022 год.

Выберите два верных утверждения.

- ☑ Целевой переменной в данной задаче является количество купленных ручек
- ☑ Решается задача регрессии
- ☐ Решается задача обучения без учителя
- ☐ Решается задача классификации
- ☐ Целевой переменной в данной задаче является клиент

Вопрос 6

Пока нет ответа

Балл: 0,32

При решении задачи классификации accuracy на тренировочных данных оказалась равна 0.7, а на тестовых - 0.65. Что можно сказать о качестве модели?

- (●) Невозможно интерпретировать качество модели, не зная количества классов в задаче и информации о доле объектов каждого класса
- ( ) Модель сильно переобучена, поэтому для снижения переобучения в этой задаче рекомендуется использовать регуляризацию
- ( ) Модель имеет низкое качество и на тренировочных, и на тестовых данных
- ( ) Модель сильно переобучена
- ( ) Модель имеет высокое качество - как на тренировочных, так и на тестовых данных

Очистить мой выбор

Вопрос 7

Пока нет ответа

Балл: 0,32

Выберите три верных утверждения про лемматизацию текстов:

- [x] Лемматизация - это приведение слова к нормальной (словарной) форме
- [ ] В результате лемматизации количество различных токенов в документе увеличивается
- [x] В результате обучения моделей на векторизованных после лемматизации текстах переобучение обычно будет ниже, чем если не делать лемматизацию
- [ ] Лемматизация - это обработка слов, в результате которой от каждого слова остается только его основа
- [x] Лемматизация нужна, чтобы снизить количество различных словоформ в текстах

Вопрос 8

Пока нет ответа

Балл: 0,32

Астрономы решают задачу предсказания длительности путешествия от Земли до различных космических объектов в световых годах. Астрономам хочется получить как можно более точный результат, при этом для них гораздо хуже, если алгоритм завысит длительность путешествия, так как тогда астрономы не успеют провести все исследования. Занижение длительности по сравнению с правильным ответом не так страшно.

Какую из метрик астрономам лучше всего использовать для оценки качества модели?

- f1-score
- f1-weighted
- MAE
- accuracy

Очистить мой выбор

Вопрос 9

Пока нет ответа

Балл: 0,32

Какая из приведенных ниже формул обладает возможностью задавать различные шаги градиентного спуска для разных весов (для различных координат вектора весов)?

Здесь w_k - значения вектора весов на k-й итерации градиентного спуска, $\nabla Q(w)$ - градиент функции потерь, v_k- вспомогательный вектор или скаляр, η, ρ - скаляры, гиперпараметры.

- $w_{k+1} = w_k - \frac{\eta}{v_k}\nabla Q(w_k)$, где $v_k = k$
- $w_{k+1} = w_k - \frac{\eta}{\sqrt{v_k+\varepsilon}}\nabla Q(w_k)$, где $v_k = v_{k-1} + (\nabla Q(w_k))^2$
- $w_{k+1} = w_k - \eta\nabla Q(w_k)$
- $w_{k+1} = w_k - \eta v_{k+1}$, где $v_{k+1} = \rho v_k + \nabla Q(w_k)$

Вопрос 10

Пока нет ответа

Балл: 0,32

Какая из перечисленных функций используют в качестве критериев информативности для построения решающих деревьев в задаче регрессии?

- Доля ошибок классификации в листе
- Среднеквадратичная ошибка
- Энтропия
- Критерий Джини

Очистить мой выбор

Вопрос 11

Пока нет ответа

Балл: 0,32

Как, основываясь на теоретических знаниях о поведении шума, смещения и разброса, должны измениться эти показатели, если к решающему лесу добавить еще одно дерево? Выберите все подходящие варианты ответа.

- [ ] Шум увеличится
- [ ] Разброс не изменится
- [ ] Смещение увеличится
- [x] Шум не изменится
- [ ] Шум уменьшится
- [ ] Разброс увеличится
- [x] Смещение не изменится
- [x] Разброс уменьшится
- [ ] Смещение уменьшится

Предыдущий элемент курса

Тренировочный вариант 1 UPDATE

Перейти на...

Следующий элемент курса

Cheat Sheet 1

Служба поддержки сайта

Вы зашли под именем Царахова Милена Викторовна (Выход)