Тестовые вопросы по MaxAbsScaler, MinMaxScaler, StandardScaler, RobustScaler

1. Какой метод масштабирования НЕ изменяет форму распределения данных?

????Странный вопрос, они все не изменяют форму графика (или так с графиками повезло), но MaxAbsScaler гарантированно не смещает нули, поэтому я думаю что это он

MaxAbsScaler просто масштабирует каждый признак result = x/|Xmax|

2. Какой метод масштабирования наиболее устойчив к выбросам?

RobustScaler

3. Какой метод масштабирования приводит данные к нулевому среднему и единичной дисперсии?

StandardScaler

4. Какой метод масштабирования использует медиану и межквартильный размах (IQR)?

RobustScaler – result=(x-median)/IQR

5. Какой метод масштабирования приводит данные к диапазону [-1, 1]?

MaxAbsScaler, если в данных есть и положительные, и отрицательные значения

MinMaxScaler, если указать feature\_range=(-1;1)

6. Какой метод масштабирования лучше всего подходит для разреженных данных?

MaxAbsScaler, потому что не смещает нули в данных

7. Если данные имеют длинный хвост и выбросы, какой метод масштабирован предпочтительнее?

RobustScaler меньше подвержен влиянию экстремальных значений/выбросов

8. Какой метод масштабирования использует минимум и максимум признака?

MinMaxScaler

Тестовые вопросы по StandardScaler

1. StandardScaler преобразует данные так, чтобы:

Среднее значение каждого признака стало 0, а стандартное отклонение std – 1, дисперсия - 1

2. StandardScaler чувствителен к выбросам?

Да, так как он использует чувствительные к выбросам стандартное отклонение и среднее значение для вычислений. **result = (x-mean)/std**

3. Какой атрибут объекта StandardScaler содержит средние значения, вычисленные для каждого признака?

scaler.mean\_

4. В каком случае рекомендуется использовать StandardScaler:

Когда данные должны быть стандартизированы для алгоритмов, чувствительных к масштабу (многие алгоритмы в машинном обучении)

5. Применение StandardScaler эквивалентно (как изменяются данные):

**result = (x-mean)/std** Данные центрируются и масштабируются до диапазона [0;1]

6. Что изменяет StandardScaler в исходных данных:

Среднее значение mean =0, стандартное отклонение std=1

7. Если данные уже имеют нулевое среднее и единичную дисперсию, применение

StandardScaler приводит к:

Минимальным изменениям из-за погрешности вычислений

x\_result=(x-mean)/std

mean=~0

std=~1

Тестовые вопросы по MaxAbsScaler

1. MaxAbsScaler масштабирует данные выполняя:

Деление каждого признака на его максимальное абсолютное значение

**x\_scaled=x/|Xmax|**

2. После применения MaxAbsScaler все значения признаков будут находиться в диапазоне:

[-1;1], если есть и отрицательные и положительные значения

[0;1], если есть только положительные значения

[-1;0], если есть только отрицательные значения

3. MaxAbsScaler хорошо подходит для данных:

Разреженных, т.к. не изменяет форму распределения данных

Когда надо сохранить нулевые элементы, знак значений и относительные расстояния между точками в признаке

4. В результате MaxAbsScaler сохраняет:

Разреженность данных, нулевые элементы, знак, относительные расстояния между точками в признаке

5. Если в данных есть только положительные значения, MaxAbsScaler будет работать

аналогично:

MinMaxScaler с feature\_range=(0, 1)

6. Что MaxAbsScaler изменяет в исходных данных:

Масштаб, нормируя данные относительно максимального абсолютного значения

7. Если максимальное абсолютное значение признака равно 1, то MaxAbsScaler:

Не изменит значений этого признака, т.к. этот признак уже масштабирован

**x\_scaled=x/|Xmax|**

**Xmax=1**

Тестовые вопросы по MinMaxScaler

1. MinMaxScaler масштабирует данные в диапазон:

[0;1] по умолчанию или указанный пользователем в feature\_range=(min,max)

2. Что MinMaxScaler использует для масштабирования:

**X\_scaled=((X-Xmin)/(Xmax-Xmin))\*(max-min)+min**

3. MinMaxScaler чувствителен к выбросам?

Да, он чувствительный к выбросам, т.к. использует чувствительные для выбросов характеристики – максимум и минимум

4. Какой атрибут объекта MinMaxScaler содержит минимальные значения, вычисленные для каждого признака?

scaled.min\_

5. Что MinMaxScaler сохраняет в исходных данных:

Отношения между значениями, относительные расстояния между точками внутри признака

6. Какой параметр MinMaxScaler позволяет задать нужный диапазон масштабирования?

feature\_range=(min,max)

7. Если данные уже находятся в диапазоне [0, 1], применение MinMaxScaler:

Не изменит данные

Тестовые вопросы по RobustScaler

1. Что RobustScaler использует для масштабирования:

X\_scaled=(X-Xmedian)/IQR=(X-Xmedian)/(Q3-Q1)

2. RobustScaler менее чувствителен к выбросам, чем:

StandardScaler, MinMaxScaler, MaxAbsScaler, т.к. использует устойчивые к выбросам медиану и межквартальный размах IQR

3. Какой параметр RobustScaler контролирует диапазон квантилей для вычисления IQR?

quantile\_range=(25,75) по умолчанию

4. ***with\_centering=False*** в RobustScaler означает:

Данные не будут центрироваться по медиане, только масштабируются по IQR, нули останутся нулями

5. RobustScaler рекомендуется использовать, когда:

Данные содержат выбросы

6. После применения RobustScaler с параметрами по умолчанию, медиана каждого признака будет:

Равна 0, т.к. данные центрируются по медиане

7. IQR (Interquartile Range) вычисляется как:

IQR=Q3-Q1, где

Q3 – третий квартиль (75-й процентиль) (25% чисел меньше этого значения)

Q1- первый квартиль (25-й процентиль) (75% чисел меньше этого значения)