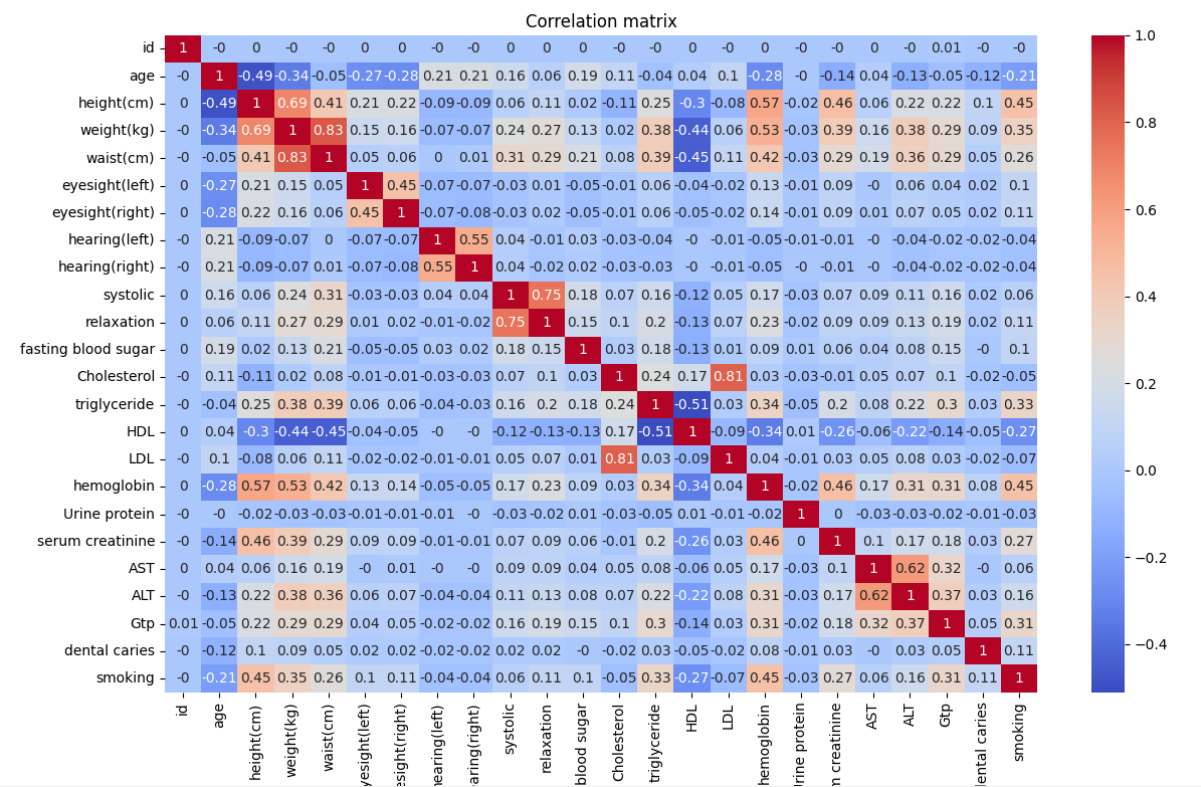


Смотрим данные

```
C:\Users\Anatoly\PycharmProjects\ML\venv\Scripts\python.exe C:\Users\Anatoly\PycharmProjects\ML\main.py
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 159256 entries, 0 to 159255
Data columns (total 24 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    159256 non-null  int64
1   age                   159256 non-null  int64
2   height(cm)           159256 non-null  int64
3   weight(kg)            159256 non-null  int64
4   waist(cm)             159256 non-null  float64
5   eyesight(left)        159256 non-null  float64
6   eyesight(right)       159256 non-null  float64
7   hearing(left)         159256 non-null  int64
8   hearing(right)        159256 non-null  int64
9   systolic              159256 non-null  int64
10  relaxation             159256 non-null  int64
11  fasting blood sugar    159256 non-null  int64
12  Cholesterol            159256 non-null  int64
13  triglyceride           159256 non-null  int64
14  HDL                   159256 non-null  int64
15  LDL                   159256 non-null  int64
16  hemoglobin            159256 non-null  float64
17  Urine protein         159256 non-null  int64
18  serum creatinine      159256 non-null  float64
19  AST                   159256 non-null  int64
20  ALT                   159256 non-null  int64
21  Gtp                   159256 non-null  int64
22  dental caries         159256 non-null  int64
23  smoking               159256 non-null  int64
dtypes: float64(5), int64(19)
memory usage: 29.2 MB
```

Обзор числовых признаков

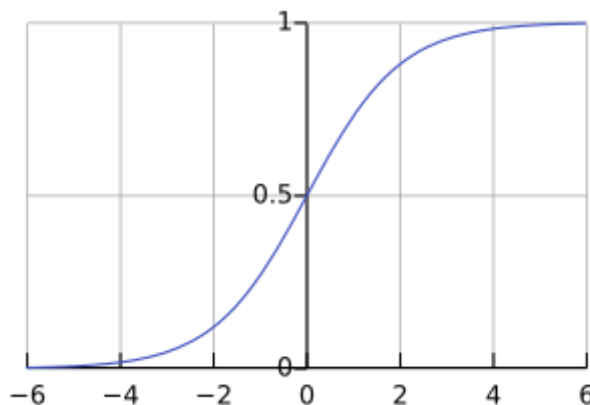


Логистическая регрессия: основные концепции и применение

Логистическая регрессия (Logistic Regression) — это алгоритм машинного обучения, используемый для задач бинарной классификации, то есть для предсказания вероятности принадлежности объекта к одному из двух классов.

Основные концепции логистической регрессии:

- **Функция логистического отклика:** В отличие от линейной регрессии, которая предсказывает непрерывные значения, логистическая регрессия предсказывает вероятность, которая ограничена диапазоном от 0 до 1. Это достигается с помощью функции логистического отклика (сигмоиды), которая преобразует линейное предсказание в вероятность:



- **Оптимизация:** Для обучения модели логистической регрессии используется метод максимального правдоподобия, который находит такие значения параметров модели, которые максимизируют вероятность наблюдаемых данных.
- **Регуляризация:** Как и в случае SVM, в логистической регрессии можно использовать регуляризацию для предотвращения переобучения. Наиболее распространенные типы регуляризации — L1 (Lasso) и L2 (Ridge).

Преимущества логистической регрессии:

- **Простота и интерпретируемость:** Логистическая регрессия — это простой и легко интерпретируемый алгоритм. Коэффициенты модели можно интерпретировать как силу и направление влияния признаков на вероятность принадлежности к классу.
- **Эффективность:** Логистическая регрессия эффективно работает с большими объемами данных и может быть легко распараллелена.
- **Гибкость:** Логистическая регрессия может быть обобщена на задачи многоклассовой классификации с использованием методов "один против всех" или "один против одного".

Недостатки логистической регрессии:

- **Предположение линейности:** Логистическая регрессия предполагает линейную зависимость между признаками и логарифмом шансов. Если это предположение нарушается, то точность модели может снизиться.

- **Чувствительность к выбросам:** Логистическая регрессия чувствительна к выбросам в данных, которые могут негативно повлиять на качество модели.

Оценка работы модели

1. Precision:
 - Значение: 0.702
 - Метрика precision измеряет, какая доля предсказанных положительных примеров действительно является положительными.
2. Recall:
 - Значение: 0.733
 - Recall измеряет, какая доля всех действительно положительных примеров была правильно предсказана моделью.
3. F1-score:
 - Значение: 0.717
 - F1-score представляет собой гармоническое среднее между precision и recall, что позволяет учитывать баланс между этими двумя метриками.
4. Матрица ошибок (confusion matrix):
 - Значение:
 - Предсказано: Положительный Предсказано: Отрицательный
 - Истинный: Положительный 12943 4840
 - Истинный: Отрицательный 2877 11192
 - Матрица ошибок показывает, насколько хорошо модель предсказывает классы, и предоставляет детальную информацию о типах ошибок, которые делает модель.
5. Точность модели (custom Accuracy):
 - Значение: 74.47%

Вывод:

- Модель правильно предсказывает значения с точностью в 74.47%, что является хорошим результатом для модели классификации