

Билет 12

1. Задача бинарной классификации.

Классификация с бинарной переменной класса, т.е. категориальной выходной переменной, которая может принимать только два значения. Очевидно, что в таких задачах решается вопрос о принадлежности объекта к одному из двух классов. Чаще всего используют состояния **0** и **1**, но могут быть и другие, например, **Да** или **Нет**, **У** или **N**, и т.д.

2. Логистическая регрессия. Алгоритм применения логистической регрессии.

Логистическая регрессия или логит-модель (англ. logit model) — статистическая модель, используемая для прогнозирования вероятности возникновения некоторого события путём его сравнения с логистической кривой. Эта регрессия выдаёт ответ в виде вероятности бинарного события (**1** или **0**).

Логистическая регрессия – это разновидность множественной регрессии, предназначенная для классификации записей на основании значений входных полей. При этом выходная переменная является категориальной или бинарной (т.е. может принимать только два значения).

В бинарной классификации каждый объект или наблюдение должны быть отнесены к одному из двух классов (например, **A** и **B**). Тогда с каждым исходом связано событие:

- Объект принадлежит к классу **A** и объект принадлежит к классу **B**.
- Результатом будет оценка вероятности соответствующего исхода.

Если в процессе анализа будет установлено, что вероятность принадлежности объекта с заданным набором значений признаков (входных переменных) к классу **A** больше, чем вероятность его принадлежности к классу **B**, то он будет классифицирован, как объект класса **A**.

Например, если рассматривается исход по займу, задается переменная **y** со значениями **1** и **0**, где **1** означает, что соответствующий заемщик расплатился по кредиту, а **0**, что имел место дефолт.

Несомненным преимуществом логистической регрессии является наличие эффективного инструмента оценки качества моделей - ROC-анализа.

Практика

```
class DataAnalysis:
    ...
    Дано вещественное число – цена 1 кг конфет.
    Вывести стоимость 1, 2, ... 10 кг конфет.
    Решить задачу используя циклическую конструкцию for.
    ...

    def __init__(self, cost=None):
        self._cost = cost
        self._costPerKilogram = {}
```

```
        self._calcCosts()

    def _calcCosts(self):
        for i in range(1, 11):
            self._costPerKilogram[i] = i * self._cost

    def showCost(self):
        for key, value in self._costPerKilogram.items():
            print(f'Стоимость за {key} кг - {value}')

task = DataAnalysis(10)
task.showCost()
```