

# Листок 03. Задача классификации

Н.В. Артамонов

29 ноября 2024 г.

## Содержание

|          |                                     |          |
|----------|-------------------------------------|----------|
| <b>1</b> | <b>Логистическая регрессия. SVM</b> | <b>1</b> |
| <b>2</b> | <b>k-NN</b>                         | <b>4</b> |
| <b>3</b> | <b>Валидация моделей</b>            | <b>5</b> |

## 1 Логистическая регрессия. SVM

#1. Для набора данных `sleep75` рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки      |
|---------------------|---------------------------|
| male                | sleep, totwrk, age, south |

1. Подгоните логистическую регрессию и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south |
|-------|-------|--------|-----|-------|
| 0     | 2900  | 2160   | 32  | 1     |
| 1     | 3120  | 1720   | 24  | 0     |
| 2     | 2850  | 2390   | 44  | 0     |

вычислите прогноз для **male**

#2. Для набора данных **sleep75** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки      |
|---------------------|---------------------------|
| male                | sleep, totwrk, age, south |

1. Подгоните модель SVM и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south |
|-------|-------|--------|-----|-------|
| 0     | 2900  | 2160   | 32  | 1     |
| 1     | 3120  | 1720   | 24  | 0     |
| 2     | 2850  | 2390   | 44  | 0     |

вычислите прогноз для **male**

#3. Для набора данных **sleep75** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки                          |
|---------------------|---|
| smsa                | sleep, totwrk, age, south, male, yngkid, marr |

1. Подгоните логистическую регрессию и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south | male | yngkid | marr |
|-------|-------|--------|-----|-------|------|--------|------|
| 0     | 2900  | 2150   | 37  | 0     | 1    | 0      | 1    |
| 1     | 3120  | 1950   | 28  | 1     | 1    | 1      | 0    |
| 2     | 2850  | 2240   | 26  | 0     | 0    | 0      | 0    |

вычислите прогноз для **smsa**

#4. Для набора данных **sleep75** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки                          |
|---------------------|---|
| smsa                | sleep, totwrk, age, south, male, yngkid, marr |

1. Подгоните модель SVM и выведите коэффициенты подогнанной модели

2. Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south | male | yngkid | marr |
|-------|-------|--------|-----|-------|------|--------|------|
| 0     | 2900  | 2150   | 37  | 0     | 1    | 0      | 1    |
| 1     | 3120  | 1950   | 28  | 1     | 1    | 1      | 0    |
| 2     | 2850  | 2240   | 26  | 0     | 0    | 0      | 0    |

вычислите прогноз для **smsa**

- #5. Для набора данных **default** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки           |
|---------------------|--------------------------------|
| default             | age, income, ownrent, selfempl |

1. Подгоните логистическую регрессию и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

| index | age  | income | ownrent | selfempl |
|-------|------|--------|---------|----------|
| 0     | 37   | 2000   | 0       | 1        |
| 1     | 42.5 | 5250   | 1       | 0        |
| 2     | 29   | 2916   | 0       | 0        |

вычислите прогноз для **default**

- #6. Для набора данных **default** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки           |
|---------------------|--------------------------------|
| default             | age, income, ownrent, selfempl |

1. Подгоните модель SVM и выведите коэффициенты подогнанной модели
2. Рассмотрим трёх людей с характеристиками

| index | age  | income | ownrent | selfempl |
|-------|------|--------|---------|----------|
| 0     | 37   | 2000   | 0       | 1        |
| 1     | 42.5 | 5250   | 1       | 0        |
| 2     | 29   | 2916   | 0       | 0        |

вычислите прогноз для **default**

## 2 k-NN

#1. Для набора данных **sleep75** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки      |
|---------------------|---------------------------|
| male                | sleep, totwrk, age, south |

Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south |
|-------|-------|--------|-----|-------|
| 0     | 2900  | 2160   | 32  | 1     |
| 1     | 3120  | 1720   | 24  | 0     |
| 2     | 2850  | 2390   | 44  | 0     |

Постройте прогноз для **male** методом k-NN с параметрами

| № | $k$ | веса     |
|---|-----|----------|
| 1 | 5   | uniform  |
| 2 | 5   | distance |
| 3 | 10  | uniform  |
| 4 | 10  | distance |

#2. Для набора данных **sleep75** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки                          |
|---------------------|---|
| smsa                | sleep, totwrk, age, south, male, yngkid, marr |

Рассмотрим трёх людей с характеристиками

| index | sleep | totwrk | age | south | male | yngkid | marr |
|-------|-------|--------|-----|-------|------|--------|------|
| 0     | 2900  | 2150   | 37  | 0     | 1    | 0      | 1    |
| 1     | 3120  | 1950   | 28  | 1     | 1    | 1      | 0    |
| 2     | 2850  | 2240   | 26  | 0     | 0    | 0      | 0    |

Постройте прогноз для **smsa** методом k-NN с параметрами

| № | $k$ | веса     |
|---|-----|----------|
| 1 | 5   | uniform  |
| 2 | 5   | distance |
| 3 | 10  | uniform  |
| 4 | 10  | distance |

#3. Для набора данных **default** рассмотрим переменные

| Зависимая/таргетная | объясняющие/признаки           |
|---------------------|--------------------------------|
| default             | age, income, ownrent, selfempl |

Рассмотрим трёх людей с характеристиками

| index | age  | income | ownrent | selfempl |
|-------|------|--------|---------|----------|
| 0     | 37   | 2000   | 0       | 1        |
| 1     | 42.5 | 5250   | 1       | 0        |
| 2     | 29   | 2916   | 0       | 0        |

Постройте прогноз для **default** методом k-NN с параметрами

| № | $k$ | веса     |
|---|-----|----------|
| 1 | 5   | uniform  |
| 2 | 5   | distance |
| 3 | 10  | uniform  |
| 4 | 10  | distance |

### 3 Валидация моделей

#1. Набор данных **sleep75** разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

| зависимая/target | объясняющая/предикторы/features |
|------------------|---------------------------------|
| male             | sleep, totwrk, age, south       |

и следующие модели

| № | Модель                            |
|---|-----------------------------------|
| 1 | Логистическая регрессия           |
| 2 | SVM                               |
| 3 | k-NN с $k = 5$ , веса 'uniform'   |
| 4 | k-NN с $k = 5$ , веса 'distance'  |
| 5 | k-NN с $k = 10$ , веса 'uniform'  |
| 6 | k-NN с $k = 10$ , веса 'distance' |

Проведите валидацию моделей относительно метрик Ассигасы, ROC. Какая модель предпочтительней?

#2. Набор данных `sleep75` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

| зависимая/target | объясняющая/предикторы/features               |
|------------------|---|
| smsa             | sleep, totwrk, age, south, male, yngkid, marr |

и следующие модели

| № | Модель                            |
|---|-----------------------------------|
| 1 | Логистическая регрессия           |
| 2 | SVM                               |
| 3 | k-NN с $k = 5$ , веса 'uniform'   |
| 4 | k-NN с $k = 5$ , веса 'distance'  |
| 5 | k-NN с $k = 10$ , веса 'uniform'  |
| 6 | k-NN с $k = 10$ , веса 'distance' |

Проведите валидацию моделей относительно метрик Ассурасы, ROC. Какая модель предпочтительней?

#3. Набор данных `default` разбейте на обучающую и тестовую часть в соотношении 80:20.

Рассмотрим задачу прогнозирования для переменных

| зависимая/target | объясняющая/предикторы/features               |
|------------------|---|
| smsa             | sleep, totwrk, age, south, male, yngkid, marr |

и следующие модели

| № | Модель                            |
|---|-----------------------------------|
| 1 | Логистическая регрессия           |
| 2 | SVM                               |
| 3 | k-NN с $k = 5$ , веса 'uniform'   |
| 4 | k-NN с $k = 5$ , веса 'distance'  |
| 5 | k-NN с $k = 10$ , веса 'uniform'  |
| 6 | k-NN с $k = 10$ , веса 'distance' |

Проведите валидацию моделей относительно метрик Ассурасы, ROC. Какая модель предпочтительней?