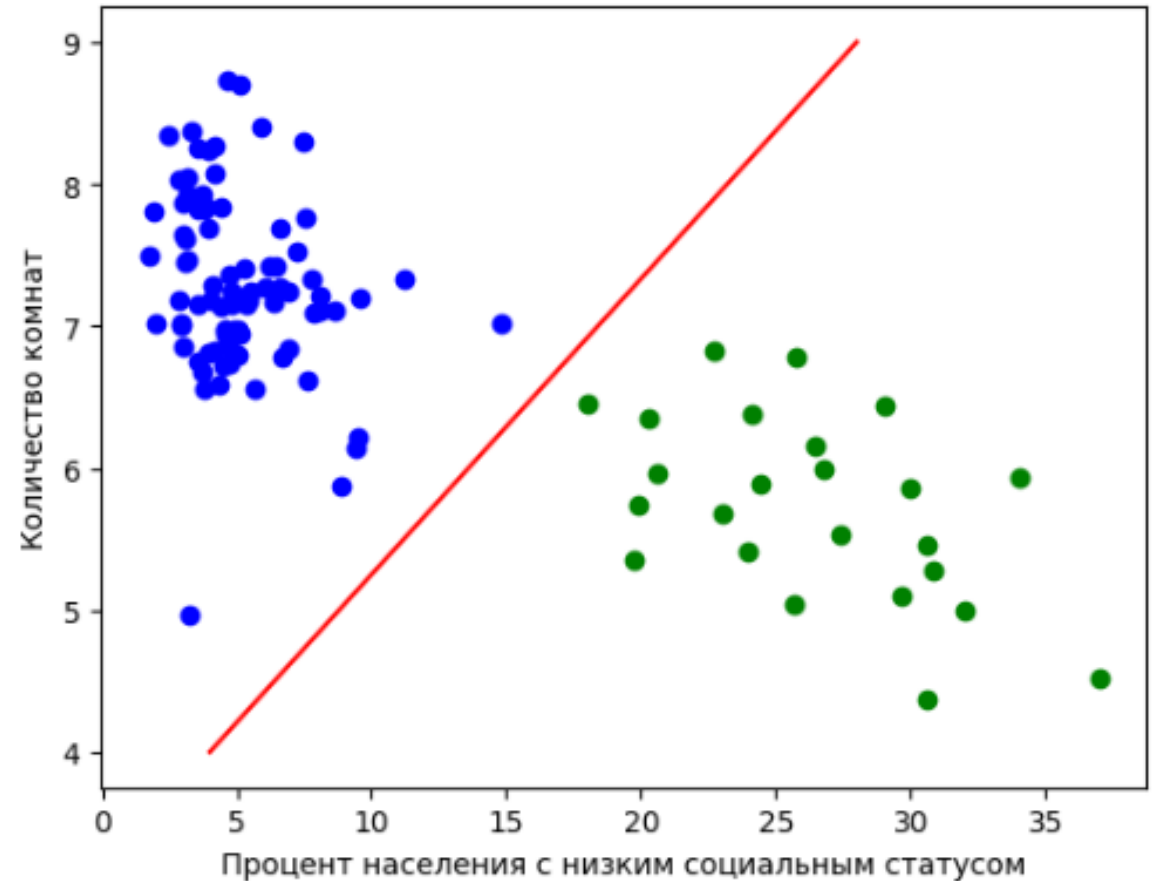


Классификация.

Логистическая регрессия

# Постановка задачи классификации

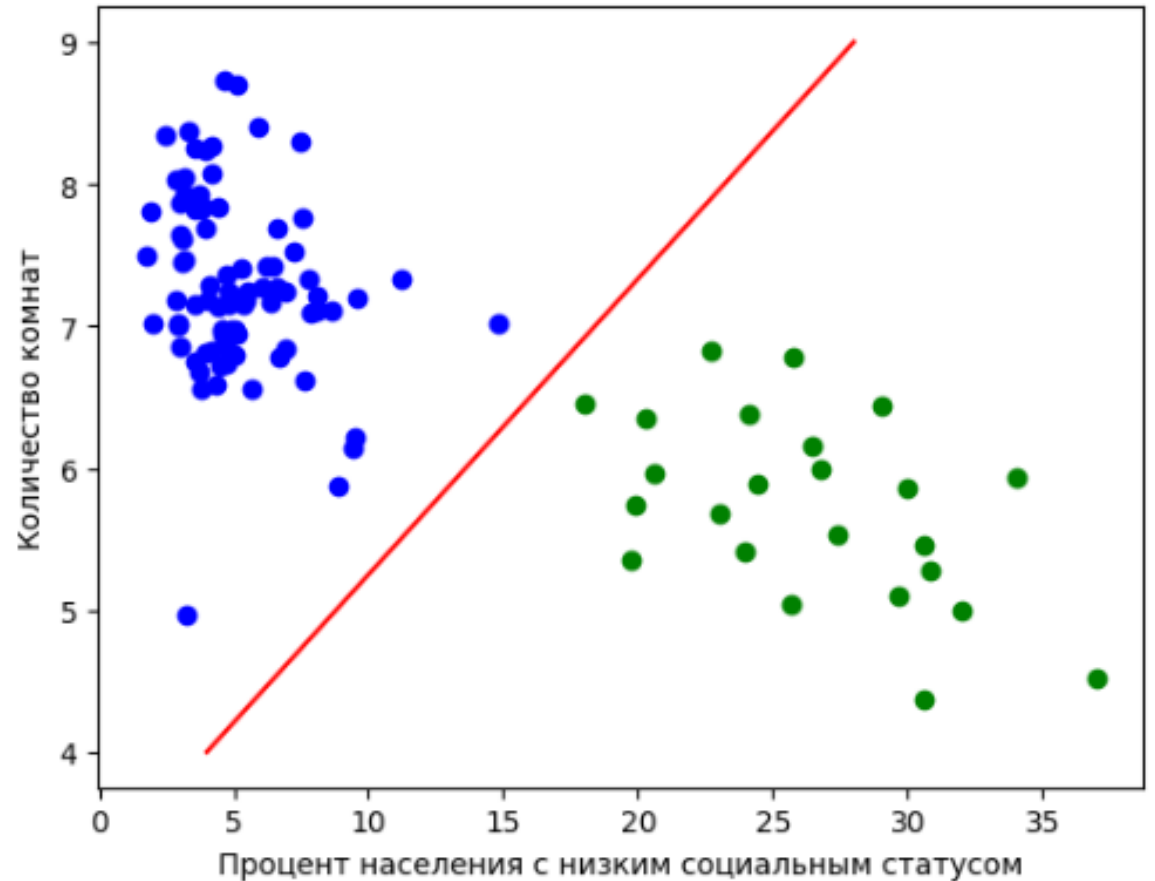
- Необходимо по признакам разделить выборку на классы
- Считается вероятность принадлежности метрики к классу



# Линейная классификация

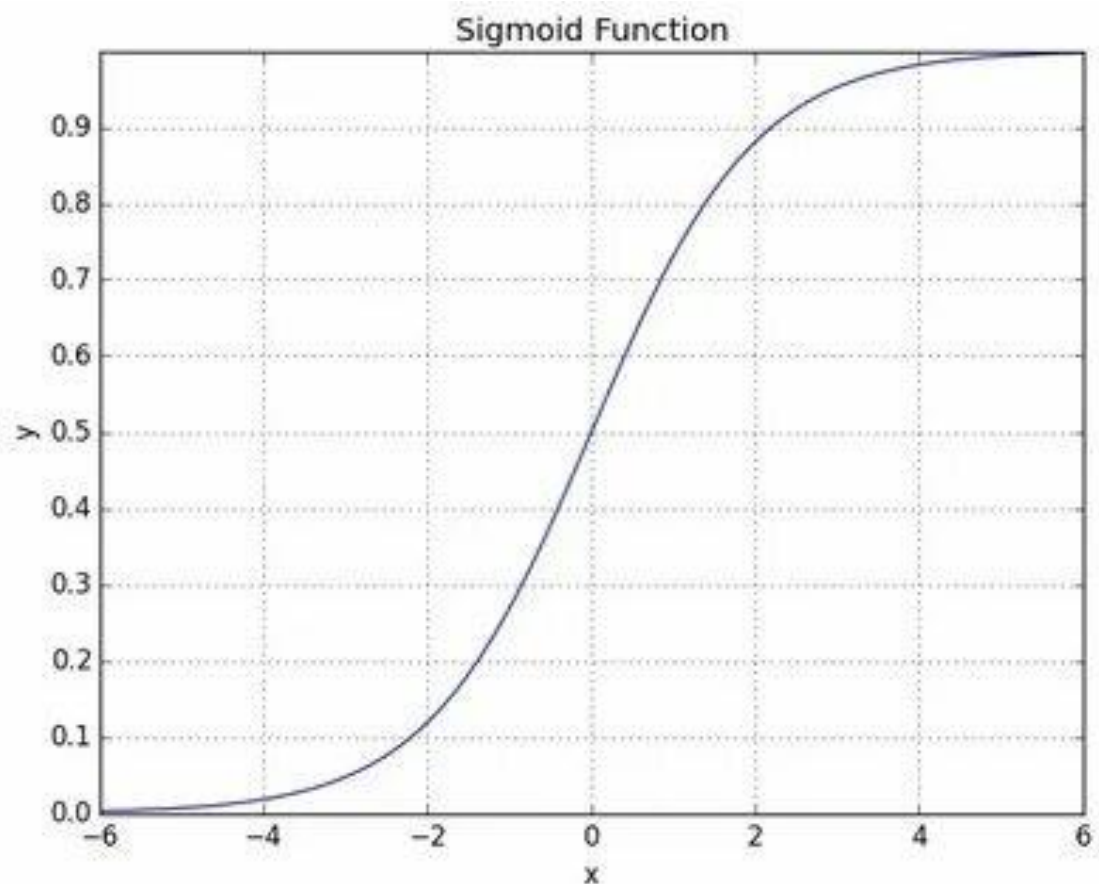
Для бинарной классификации:

- $\hat{y} = \text{sign}(z)$
- $z = w_0 + w_1 x_1 + \dots + w_k x_k$ ;
- $z = \langle x, w \rangle$ ,  $x = (1, x_1, \dots, x_k)$



# Логистическая регрессия

- $\hat{y} = \frac{1}{1+e^{-z}} = \sigma(z)$ ;
- Значения в диапазоне  $[0,1]$
- $\frac{1}{2}$  в нуле
- Моделирует вероятность принадлежности к классу
- В общем случае:  
 $\hat{y} = \sigma(Xw)$



# Логистическая регрессия, обучение

- Применяется функция потерь

$$\ln(1 + e^{-yz}) \rightarrow \min$$

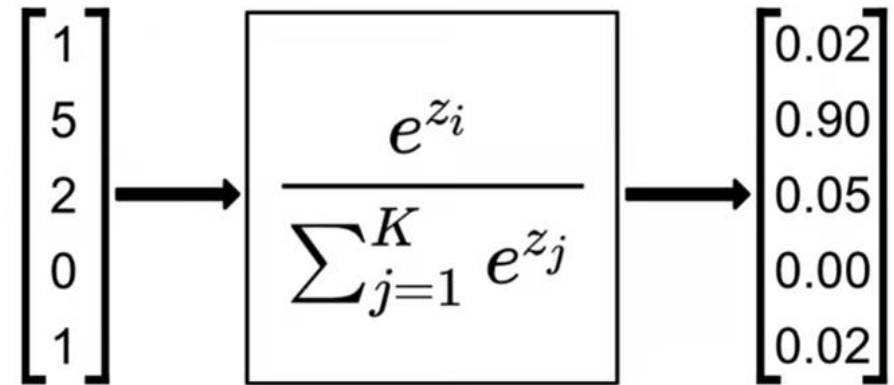
по всем  $w$

- Максимизируют логарифм правдоподобия (обычно методом Ньютона)

$$L = \sum_{i=1}^N \ln(\hat{y}_i) \rightarrow \max$$

# Многоклассовая классификация

- Может быть использована, для этого обучают  $K$  моделей, каждая из которых отличается только целевым классом и выдается совокупное решение нескольких моделей  $\text{softmax}(x)$ .
- Другими словами, решается задача бинарной классификации  $K$  раз и применяется усреднение



# Логистическая регрессия

- Линейный алгоритм классификации
- Вычисляет вероятность принадлежности к классам
- Перед использованием лучше провести нормализацию данных

# Многоклассовая классификация

Если у принимает больше двух значений, то можно использовать метод «все против всех»:

- Для каждой пары возможных значений, составить выборку, содержащую только эти значения
- На каждой из этих выборок запустить алгоритм
- В результате мы получили набор моделей, голосующих каждая за свой прогноз. Выбираем прогноз большинством голосов