

Линейная и логистическая регрессии

Наталья Баданина

Методист академических программ ML направления, Яндекс

Data Scientist



Проверка связи





Если у вас нет звука:

- убедитесь, что на вашем устройстве и колонках включён звук
- обновите страницу вебинара или закройте её и заново присоединитесь к вебинару
- откройте вебинар в другом браузере
- перезагрузите устройство и попытайтесь зайти заново



Поставьте в чат:

-  если меня видно и слышно
-  если нет

Рекомендации

→ При просмотре с компьютера

- Используйте браузеры **Google Chrome** или **Microsoft Edge**
- Если есть проблемы с изображением или звуком, обновите страницу — **F5**

→ При просмотре с мобильного телефона или планшета

- Перейдите с мобильного интернет-соединения на **Wi-Fi**
- Если есть проблемы с изображением или звуком, перезапустите приложение на телефоне

Правила участия

- 1 Приготовьте блокнот и ручку, чтобы записывать важные мысли и идеи
- 2 Продолжительность вебинара — 90 минут
- 3 Вы можете писать свои вопросы в чате
- 4 Запись вебинара будет доступна в личном кабинете



Наталья Баданина

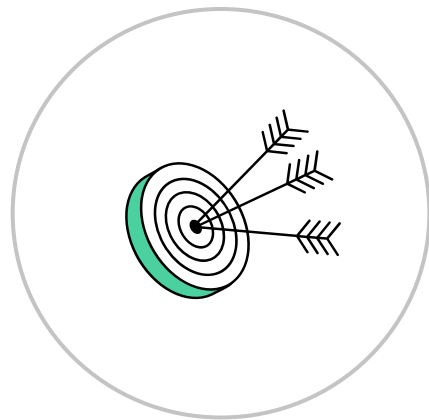
О спикере:

- Методист академических программ ML направления, Яндекс
- Data Scientist



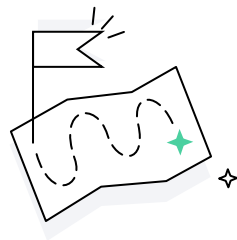
Цели занятия

- 1 Познакомиться с понятием линейной регрессии
- 2 Познакомиться с понятием линейной регрессии логистической регрессии
- 3 Применить на практике модель логистической регрессии и вычислить основные метрики



План занятия

- 1 Линейная регрессия
- 2 Логистическая регрессия
- 3 Практика



Линейная регрессия



1

Линейная регрессия

Гипотеза о линейной зависимости целевой переменной, ищем решение в виде:

$$a(X_1, \dots, X_n) = w_0 + w_1 X_1 + \dots + w_n X_n$$

- *часто неплохо работает и при монотонных зависимостях*
- *хорошо работает, когда есть много «однородных» зависимостей*

Линейная регрессия от одной переменной

$$a(X_1) = w_0 + w_1 X_1$$

хотели бы

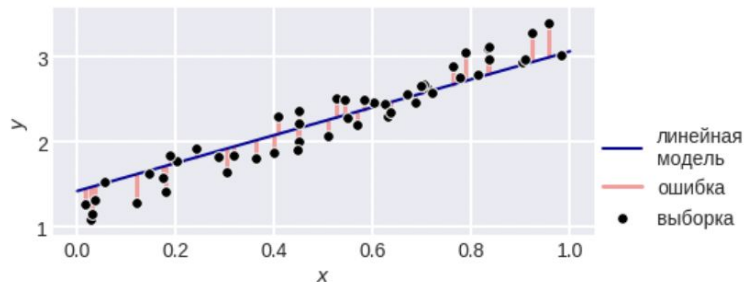
$$\begin{cases} w_0 + w_1 x_1 = y_1 \\ \dots \\ w_0 + w_1 x_m = y_m \end{cases}$$

невязки/отклонения (residuals)

$$e_1 = y_1 - w_0 - w_1 x_1$$

...

$$e_m = y_m - w_0 - w_1 x_m$$



Линейная регрессия от одной переменной

Задача минимизации суммы квадратов отклонений
(residual sum of squares):

$$\text{RSS} = e_1^2 + \dots + e_m^2 \rightarrow \min$$

На это можно смотреть как на минимизацию эмпирического риска по параметрам $w = (w_0, w_1)$

$$L(w) = \sum_{i=1}^m (y_i - a_w(x_i))^2 = \sum_{i=1}^m (y_i - (w_0 + w_1 x_i))^2$$



Ваши вопросы?

Логистическая регрессия



2

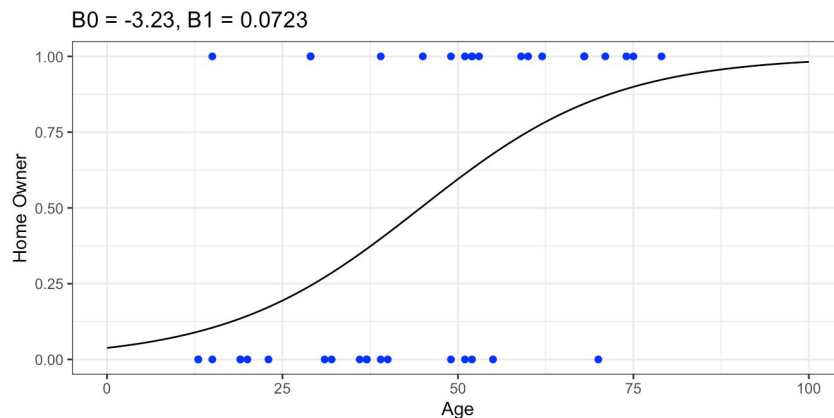
Логистическая регрессия

Решаем задачу классификации методом логистической регрессии

$$p(x) \equiv P(Y = 1 | x) = \sigma(z) = \frac{1}{1 + e^{-z}} \in (0, 1),$$

$$z = w^T x = w_0 + w_1 X_1 + \dots + w_n X_n,$$

$$\log\left(\frac{p(x)}{1 - p(x)}\right) = z$$



Практика



3

Цель и задачи

Цель:

1. Изучить практическую реализацию логистической регрессии

Задачи:

1. Понять, как обучить модель
2. Разобрать теоретическую составляющую модели
3. Изучить метрики классификации для оценки модели



Ваши вопросы?

Выводы

1. Линейная и логистическая регрессии связаны, но решают разные задачи – регрессия и классификация
2. Для оценки модели нужно использовать несколько метрик

Итоги занятия

- Познакомились с понятием линейной регрессии
- Познакомились с понятием логистической регрессии
- Применили на практике модель логистической регрессии и вычислили основные метрики



Рефлексия

- Что изменилось? Раньше я думал(а), что..., а теперь...
- Какие вопросы у меня остались?



Домашнее задание. ДЗ 2. Часть 2

1. Проведите встречу команды и распределите задачи, используя материалы Вебинара «Линейная и логистическая регрессии»
2. Реализуйте задачи:
 - Примените модель, соответствующую вашей задаче
 - Посчитайте метрики

Срок выполнения Части 1 и Части 2 задания — 7 дней с момента открытия задания

Линейная и логистическая регрессии

Наталья Баданина

Методист академических программ ML направления, Яндекс

Data Scientist

