

Линейный классификатор и логистическая регрессия

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ

Вопросы занятия

1. Линейные модели: требования к данным и практика;
2. Логистическая регрессия: практическое задание.

В конце занятия научимся:

- изучим преимущества и недостатки линейных моделей, а также требования к данным;
- реализовывать алгоритм линейной и логистической регрессии.

Линейные модели

ОСНОВНЫЕ ДОСТОИНСТВА

- Линейные модели подходят для описания многих процессов
- Относительная простота вычислений и интерпретации результатов
- Вклад нескольких факторов часто можно разбить на сумму влияния каждого фактора в отдельности

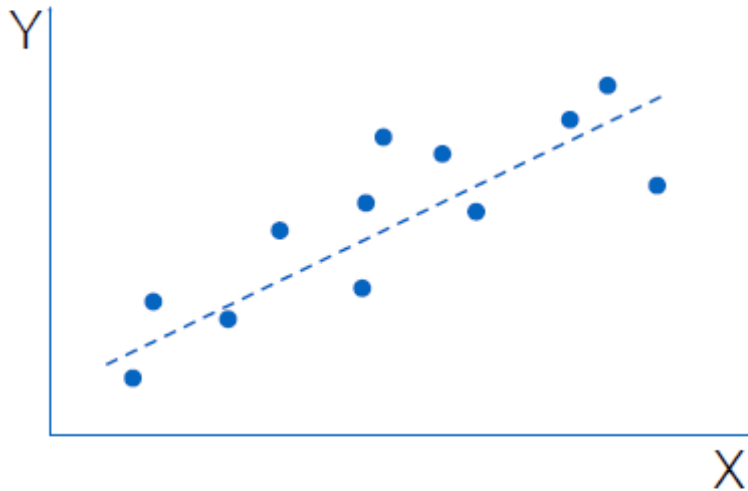
Линейные модели

ОСНОВНЫЕ ДОСТОИНСТВА

- Прогноз продаж по объему инвентаря, загрузке, площади и другим «линейным» характеристикам
- Построение вероятностных моделей в страховании, кредитном скоринге, инвестиционных проектах
- Предсказание цены товара на основании его характеристик
- Построение трендов

Линейные модели

ОПРЕДЕЛЕНИЕ



$$y_i = \sum_{j=1}^m w_j X_{ij} + e_i$$

Y – целевая переменная

W – вектор весов модели

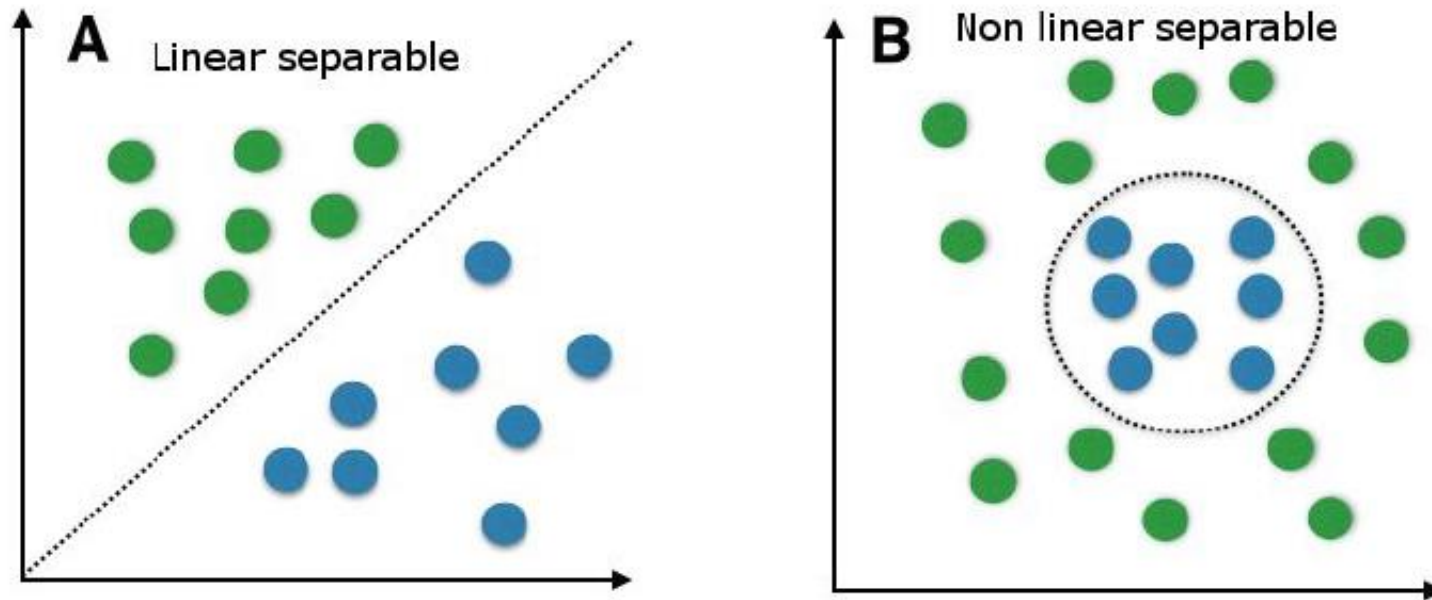
X – матрица наблюдений

e – ошибка модели

Линейная регрессия (Linear regression) — модель зависимости переменной x от одной или нескольких других переменных (факторов, регрессоров, независимых переменных) с линейной функцией зависимости.

Построение линейной модели

КАК СТРОИМ ЛИНЕЙНУЮ МОДЕЛЬ

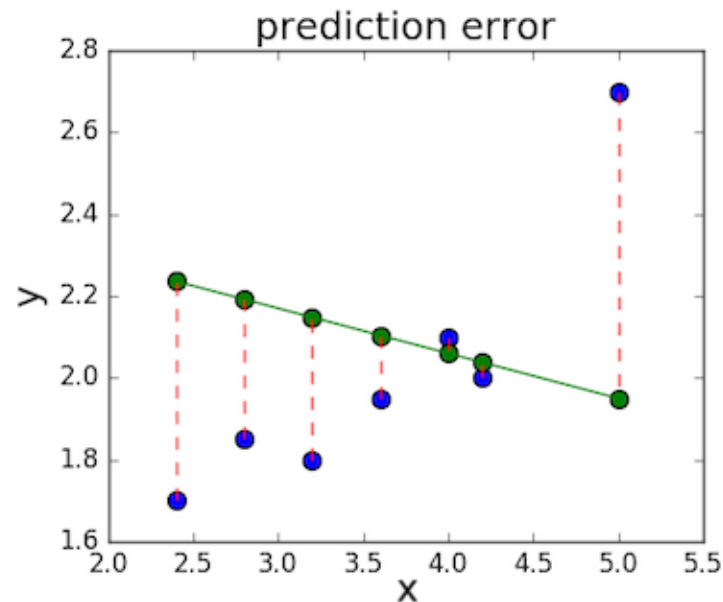


Один из способов вычислить значения параметров модели является *метод наименьших квадратов*, который минимизирует среднеквадратичную ошибку между реальным значением зависимой переменной и прогнозом, выданным моделью.

Построение линейной модели

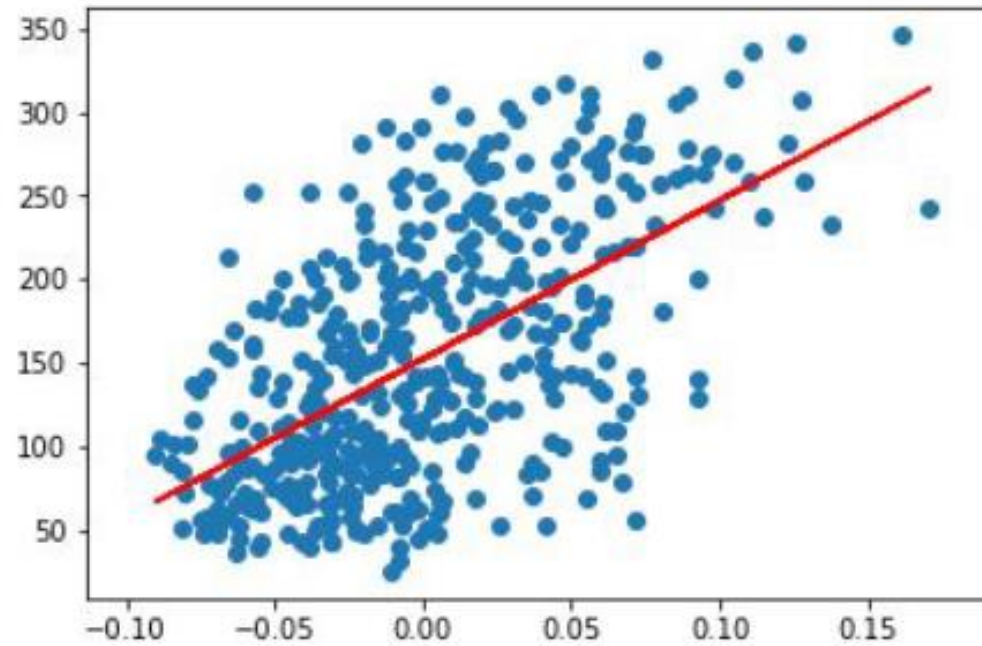
Функция потерь — это мера количества ошибок, которые наша линейная регрессия делает на наборе данных. Хотя есть разные функции потерь, все они вычисляют расстояние между предсказанным значением $y(x)$ и его фактическим значением.

Наиболее распространенная функция потерь называется средней квадратичной ошибкой (MSE). Чтобы вычислить MSE, мы просто берем все значения ошибок, считаем их квадраты длин и усредняем.



Построение линейной модели

Пример из кода `Linear regression.ipynb`



ПРАКТИКА

SAT_model.IPYNB

ПРАКТИКА

REGRESSION_CARS.IPYNB

ПРАКТИКА

Регуляризация. IPYNB

ПРОГНОЗ ВЕРОЯТНОСТИ

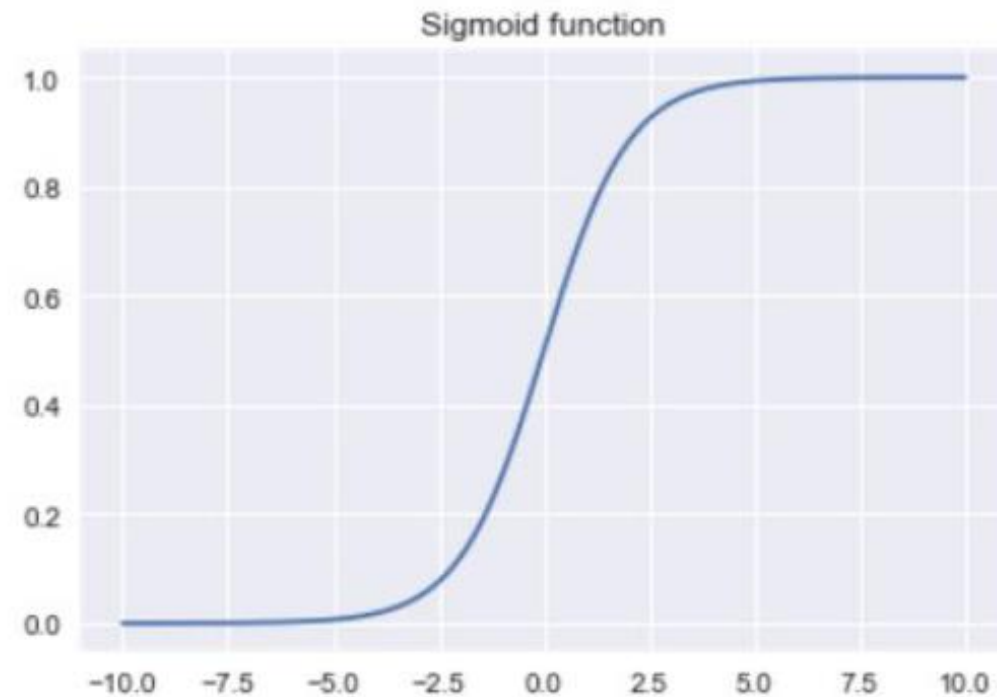
Прогнозирует вероятность отнесения наблюдения к определенному классу

Модель:
$$L = a_0 + a_1X_1 + a_2X_2 + \dots + a_nX_n$$

Логистическая регрессия

ПРОГНОЗ ВЕРОЯТНОСТИ

Вероятность: $p = \frac{1}{1 + e^{-L}}$



ПРАКТИКА

Logistic_regression_athletes_classifier.ipynb
Ознакомление и доработка

ЕСЛИ КЛАССОВ БОЛЬШЕ ДВУХ?

IRIS_DATASET.IPYNB
ознакомление

ЧТО МЫ СЕГОДНЯ УЗНАЛИ

1. Изучили линейные модели и требования к ним.
2. Реализовали логистическую регрессию.

Линейный классификатор и логистическая регрессия

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ