

Лекция 8

Введение в машинное обучение

Курс: Введение в DS на УБ и МиРА (весна, 2022)

Преподаватель: Владимир Омелюсик

16 мая 2022 г.

Что такое машинное обучение?

Неформальное определение

Машинное обучение – дисциплина, изучающая построение моделей, позволяющих компьютерам воспроизводить зависимости между разными объектами без их непосредственного программирования.

- Есть конечная выборка, на которой обучаем модель.
- В ходе обучения происходит «запоминание» зависимостей.
- После обучения модель способна давать «хорошие» предсказания на новых данных.

- Иногда можно получить явный математический вид.
- Иногда нет.
 - Какая завтра погода?
 - Какая тональность у текста?
 - На фотографии кошка или собака?
- Найти точные математические функции для ответа на эти вопросы сложно или невозможно. Но если у нас есть некоторый набор данных, то можно попытаться приблизить истинные зависимости некоторыми математическими моделями.
- Статистика – про объяснения, машинное обучение – про предсказания.

- Наблюдение:
- Признак:
- Целевая переменная:
- Модель:
- Параметры:
- Гиперпараметры:
- Обучающая выборка:
- Тестовая выборка:
- Функция потерь:
- Метрика качества:

Пример: предсказание стоимости квартиры

1. Обучение с учителем.

- Регрессия.
- Классификация.
 - Бинарная.
 - Многоклассовая.
 - С пересекающимися классами.
- Ранжирование.

2. Обучение без учителя.

- Кластеризация.
- Понижение размерности.
- Визуализация.

Задача регрессии

- $y_i \in \mathbb{R}$

Задача классификации (бинарная)

- $y_i \in \{-1, 1\}$

Задача классификации (многоклассовая)

- $y_i \in \{1, \dots, K\}$

Задача классификации (с пересекающимися классами)

- $y_i \in 0, 1^K$
- Ответ – вектор из нулей и единиц длины K .
- Единица на позиции i означает, что объект принадлежит к классу i .

Задача ранжирования

- Есть набор документов d_1, \dots, d_N и некоторый запрос q .
- Хотим сортировать документы в соответствии с релевантностью запросу.
- Алгоритм должен выдавать оценку релевантности.

Задача кластеризации

- Есть только X , а y отсутствует.
- Хотим найти группы «похожих» объектов в X , используя только характеристики X .
- Как определить «похожесть»? Как оценить качество? Как выбрать число групп?

Задача понижения размерности

- X имеет размеры $N \times d$, где d – очень большое.
- Пример: медицинские измерения.
- Проблемы:
 - Модели долго обучаются.
 - Некоторые модели могут неправильно обучиться.
- Решение – построить алгоритм, который на основании выборки X построит новую выборку с меньшим числом признаков.

Задача визуализации

- Частный случай задачи понижения размерности, где новая матрица состоит из 2 или 3 признаков.

- Всё то же самое, что обсуждали до этого.
- Важно только качество предсказаний.
- Проблемы с обучением по формулам:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

- Если матрица $(X^T X)$ необратима, то будут проблемы с вычислениями.
- Произведение матриц – долгая операция.

- В многомерном случае рассчитываем градиент:

$$\nabla_x f(x) = \left(\frac{df}{dx_1}, \dots, \frac{df}{dx_d} \right)$$

- Например, градиент MSE:

$$\nabla_{\beta} MSE = \frac{2}{N} X^T (Xw - y)$$

- Градиентный спуск для обучения:

$$\beta_{t+1} = \beta_t - \alpha \nabla_{\beta} MSE(\beta_t),$$

где $\alpha > 0$ – длина шага.

Алгоритм градиентного спуска

1. Выбираем начальное приближение β_0 .
2. Повторяем

$$\beta_{t+1} = \beta_t - \alpha \nabla_{\beta} \text{MSE}(\beta_t),$$

3. Останавливаемся, если

$$\|w_t - w_{t-1}\|_2 \leq \varepsilon$$

Проблема градиентного спуска

- Градиентный спуск находит только локальные минимумы.
- Решение: муьтистарт

$$\beta_{t+1} = \beta_t - \alpha \nabla_{\beta} MSE(\beta_t),$$

- Позволяет контролировать скорость обучения.
- Если сделать слишком большой, можно «перепрыгнуть» минимум.
- Гиперпараметр, нужно подбирать.

