

Разработка программного обеспечения на языке Python

[Обзорная панель](#) [Мои курсы](#) [Разработка ПО на языке Python](#) [Анализ данных и машинное обучение](#)

[Лекция 2. Введение в машинное обучение](#)

Лекция 2. Введение в машинное обучение

Посмотрите видеоуроки и ответьте на контрольные вопросы после лекции

Определения и понятия

Машинное обучение - это наука, изучающая способы извлечения закономерностей из ограниченного количества примеров. Есть ещё много близких направлений. Например, к анализу данных можно отнести любую работу, связанную с извлечением полезной информации из данных, даже если это извлечение делается методом пристального взгляда. Последнее время нередко говорят про искусственный интеллект, но зачастую это оказывается лишь более красивым названием для результатов, полученных методами машинного обучения. Впрочем, обсуждение терминологии вряд ли нам сильно поможет, поэтому лучше перейдём к конкретному примеру.

Представим, что нам принадлежит большая сеть аптек, и мы хотим открыть в некоем городе новую аптеку. Мы нашли несколько точек в городе, где есть возможность приобрести помещение и организовать там аптеку. Нам важно, чтобы через определенное время она стала приносить прибыль - точнее, хочется открыть её в той точке, в которой прибыль окажется максимальной. **Поставим задачу:** для каждого возможного размещения аптеки предсказать прибыль, которую она принесет в течение первого года. Объектом мы будем называть то, для чего хотим сделать предсказание. В нашем случае это конкретная точка размещения аптеки. Обозначать объект мы будем маленькой буквой x , а если их несколько, то будем добавлять нижние индексы. Множество всех возможных точек размещения называется пространством объектов и обозначается через X . Величина, которую мы хотим определять (т.е. прибыль аптеки), называется ответом или целевой переменной, а множество ее значений пространством ответов Y . В нашем случае пространство ответов является множеством вещественных чисел: $Y = \mathbb{R}$. Отдельные ответы будем обозначать маленькой буквой y .

Такой **прогноз** будем делать на основе своих экспертных знаний. У нас есть лишь **примеры** - поскольку мы владеем целой сетью аптек, то имеем данные по достаточно большому числу ранее открытых аптек и по их прибыли в течение первого года. Каждый такой пример называется обучающим, а вся их совокупность - обучающей выборкой, которая обозначается как $X = \{(x_1, y_1), \dots, (x_\ell, y_\ell)\}$, где x_1, \dots, x_ℓ - обучающие объекты, а ℓ - их количество. Особенность обучающих объектов состоит в том, что для них известны ответы y_1, \dots, y_ℓ .

Отметим, что **объекты** - это некие абстрактные сущности (точки размещения аптек), которыми компьютеры не умеют оперировать напрямую. Для дальнейшего анализа нам понадобится описать объекты с помощью некоторого набора характеристик, которые называются признаками (или факторами). Вектор всех признаков объекта x называется признаковым описанием этого объекта. Далее мы будем отождествлять объект и его признаковое описание. Признаки могут быть очень разными: бинарными, вещественными, категориальными (принимают значения из неупорядоченного множества), ординальными (принимают значения из упорядоченного множества), многозначными (set-valued, значения являются подмножествами некоторого универсального множества). Признаки могут иметь сложную внутреннюю структуру: так, в качестве признака для конкретного человека в задаче предсказания его годового дохода может служить фотография. Разумеется, фотографию можно представить и как некоторое количество бинарных или вещественных признаков, каждый из которых кодирует соответствующий пиксель изображения. Однако, работа с изображением как с одной сложной структурой позволяет вычислять по нему различные фильтры, накладывать требование инвариантности ответа к сдвигам и т.д. На работе со сложными данными специализируется активно развивающееся сейчас глубинное обучение (deep learning).

Вопросы

[◀ Задание 1. Создание датафрейма](#)

Перейти на...

СЛЕДУЮЩИЙ ЭЛЕМЕНТ КУРСА

[Анализ данных - Практическое занятие 2 ▶](#)

© 2010-2023 Центр обучающих систем
Сибирского федерального университета, sfu-kras.ru

Разработано на платформе moodle
Beta-version (3.9.1.5.w3)

[Политика конфиденциальности](#)

[Соглашение о Персональных данных](#)

[Политика допустимого использования](#)

Контакты +7(391) 206-27-05
info-ms@sfu-kras.ru

[Скачать мобильное приложение](#)

[Инструкции по работе в системе](#)