Урок 27. Обучение без учителя и кластеризация

7 класс · Урок 27 из 34 · 45 минут

Обучение без учителя и кластеризация: урок для 7 класса о поиске групп, выборе признаков, диаграмме рассеяния и смысле найденных различий.

Алгоритм нашёл похожие объекты. Кто решил, что считать похожим, и что означают получившиеся группы? Проверим это на измерениях знакомых предметов и своей диаграмме.

▼ Содержание урока

1. Цель урока
2. Ход урока
3. Этап 1. Вход в тему: сколько групп
4. Этап 2. Чем поиск групп отличается от узнавания класса
5. Этап 3. Сначала сигнал, потом объяснение
6. Этап 4. Измеряем, строим и интерпретируем
7. Этап 5. Итоги урока

▼ 01. Цель урока

Научиться искать и объяснять группы в данных, когда готовые метки классов не заданы.

К концу урока ученики смогут:

- Различать классификацию и кластеризацию.
- Объяснять зависимость групп от признаков и меры сходства.
- Создавать таблицу измерений и диаграмму рассеяния.
- Отделять наблюдаемые скопления от объяснения их причин.

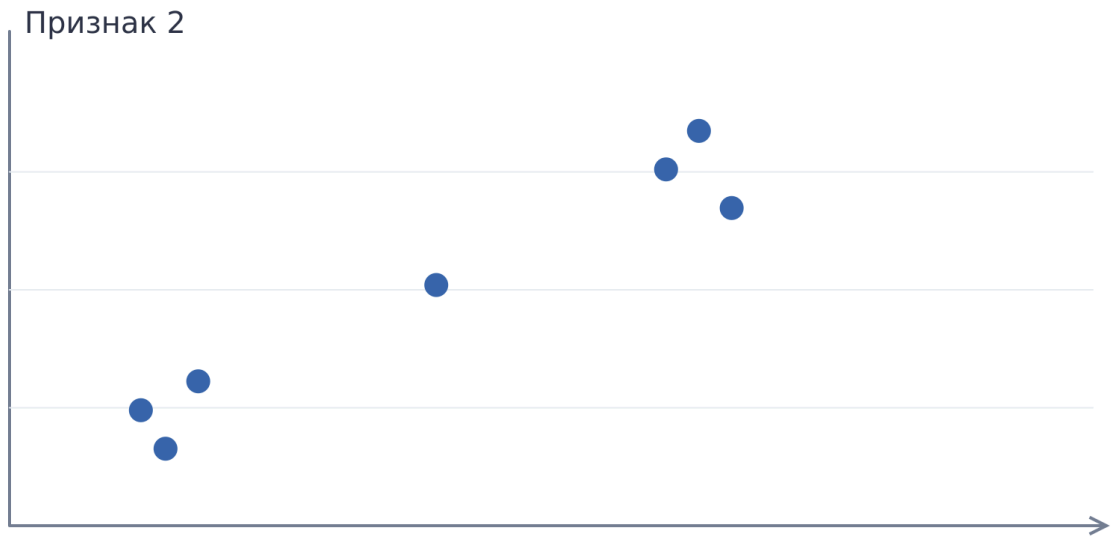
▼ 02. Ход урока

Этап	Мин	Учитель	Ученики
1. Вход в тему: сколько групп	5	Сравнивает два обоснованных способа разделения точек.	Предлагают группы и обсуждают промежуточный объект.
2. Чем поиск групп отличается от узнавания класса	10	Объясняет кластеры, признаки и масштаб.	Различают поиск групп и предсказание готового класса.
3. Сначала сигнал, потом объяснение	9	Разбирает поиск групп сейсмических записей.	Отделяют обнаружение сходства от его объяснения.
4. Измеряем, строим и интерпретируем	15	Помогает организовать измерения и построение графика.	Создают таблицу и диаграмму, обсуждают смысл групп.
5. Итоги урока	6	Обсуждает общие выводы и ответы на итоговые вопросы.	Формулируют знания и объясняют ответы.

03. Этап 1. Вход в тему: сколько групп 5 минут

Говорим ученикам

Посмотрите на схему ниже. Каждый кружок — объект с двумя числовыми признаками. Сначала самостоятельно предложите, сколько групп здесь разумно выделить. Затем сравните с соседом: куда попал промежуточный кружок?



Где вы проведёте границы?

Схема для обсуждения, а не результаты измерения. Готовые группы не выделены. Оси пока не имеют предметного смысла: его нужно задать, прежде чем объяснять группы.

Одни и те же точки можно объединять по-разному, если по-разному понимать близость и допустимый размер группы. Сегодня разберём обучение без учителя и кластеризацию — поиск похожих объектов без заранее выданных правильных меток. Обоснуйте своё разделение; допустимы разные варианты.

04. Этап 2. Чем поиск групп отличается от узнавания класса

10 минут

Говорим ученикам

В классификации известны категории и обучающие примеры с метками. В кластеризации алгоритм ищет группы по сходству самих данных: готового ответа для каждой строки нет. Найденные группы называют кластерами. Это один из видов задач обучения без учителя.

Сходство задают через признаки и способ сравнения. Если группировать музыкальные записи по темпу, получатся одни соседи, а по составу инструментов — другие. Близость зависит от того, какие признаки мы сравниваем. Человек выбирает представление, настройки и затем выясняет, полезен ли результат.

Числовые шкалы тоже влияют на сравнение. Если длину в сантиметрах заменить миллиметрами только на одной оси, её числовые различия увеличатся в десять раз. Алгоритм, который складывает расстояния по осям, может начать сильнее учитывать эту ось. Поэтому признаки приводят к сопоставимым масштабам или осознанно задают их вклад.

Название группы появляется после анализа, а не доказывается цветом точек. Два близких объекта могут оказаться рядом из-за выбранного признака; из этого ещё не следует одинаковая причина, качество или назначение. И обучение без учителя не означает, что человек вообще не участвует в работе.

Для учителя

Короткая проверка: готовые категории «учебник/тетрадь» и обучение по подписанным обложкам — классификация. Поиск групп по длине и ширине без этих меток — кластеризация. Не требуйте вводить k-means и вычислять центры: попросите учеников объяснить, как выбор признаков повлиял на получившиеся группы.

05. Этап 3. Сначала сигнал, потом объяснение 9 минут

Говорим ученикам

Исследователю часто мешают разные источники сигнала, которые наложились друг на друга. Но названия источников не всегда можно заранее подписать для каждой секунды записи. Тогда поиск похожих участков помогает навести порядок в большом потоке.

8 июля 2020 года Кристофер Джонсон и соавторы опубликовали исследование сейсмических записей — измерений слабых движений грунта. Они группировали участки по временным и частотным признакам, чтобы выделять разные виды шумовых сигналов. Это помогало отделять их при анализе слабых колебаний.

У такой работы два шага: найти сходные записи и понять, что означает их сходство. Чтобы объяснить найденную группу, нужно исследовать, почему её объекты похожи.

Реальный случай: похожие колебания среди сейсмического шума

В исследовании Джонсона и коллег признаки записей движений грунта использовали для кластеризации шумовых сигналов. Группы помогали анализировать сложный фон, среди которого ищут небольшие события.

Вывод: кластеризация может обнаружить структуру до появления полного набора меток, но смысл найденных групп нужно исследовать отдельно.

06. Этап 4. Измеряем, строим и интерпретируем 15 минут

Говорим ученикам

Для диаграммы рассеяния нужны два числовых признака одного наблюдения. По горизонтали откладываем один, по вертикали другой. Одна точка сохраняет пару значений. Если просто отсортировать один столбец отдельно от второго, получим точки несуществующих объектов.

Сейчас используем настоящие размеры учебных материалов на парте. Такие сведения пригодятся, например, при подборе обложек или разделителей. Но назначение вещи и её размер не одно и то же: это проверим по графику.

Задание. Есть ли группы по размеру?

В группе выберите 6–8 доступных книг, тетрадей или альбомов. Ничего покупать или приносить специально не нужно. Измерьте длинную и короткую стороны обложки в сантиметрах; каждой вещи дайте условный номер.

1. Создайте таблицу: номер, короткая сторона, длинная сторона. Одна строка относится к одному предмету.
2. Постройте диаграмму рассеяния в редакторе или на координатной сетке. Подпишите оси и единицы.
3. Предложите разбиение на группы или объясните, почему явных скоплений нет. Отметьте пограничные точки.
4. Сопоставьте группы с задачей подбора обложек: какие размеры ещё нужно узнать, прежде чем выбирать конкретную обложку?
5. Покажите график соседней группе. Пусть она предложит иное основание сходства; объясните, какие данные потребуются для него.

Критерии: пары измерений сохранены; график соответствует таблице; группы обоснованы; отсутствие групп допустимо; из размера не сделан вывод о качестве учебного материала.

Для учителя

Сантиметровой точности достаточно для понимания графика; это не заказ обложек по точному размеру. Для реального подбора важны также толщина, конструкция и допуски. Если предметы одинакового формата, точки могут совпасть — подпишите номера рядом, не раздвигайте их как разные измерения. Не требуйте обязательно получить два кластера. Перевод только одной оси в миллиметры можно обсудить мысленно, не перестраивая все данные. В заключение сравните график с реальными предметами.

07. Этап 5. Итоги урока 6 минут

Говорим ученикам

Кластеризация помогает увидеть различия, которые заранее не были размечены. Результат зависит от того, какое сходство мы искали. Чтобы перейти от групп к полезному решению, нужно проверить признаки, масштаб и смысл групп в своей задаче.

Главное за урок

- Кластеризация ищет группы по сходству, а классификация предсказывает заранее определённые категории.
- Группы зависят от выбранных признаков, шкал и способа сравнения.
- Диаграмма рассеяния показывает пары числовых признаков каждого объекта.
- Если на графике собрались похожие объекты, разберитесь, по каким признакам они оказались рядом. Причину сходства нужно выяснять отдельно.

Итоговые вопросы

Ответьте коротко. Затем обсудим объяснения двух-трёх участников.

1. Чем кластеризация отличается от классификации, если на обеих картинках объекты раскрашены по группам?
2. Почему изменение единицы одного признака может повлиять на результат группировки?
3. Можно ли по кластеру похожих по размеру книг сделать вывод об их одинаковой трудности?
4. Что означает отсутствие заметных скоплений на диаграмме?

Для учителя

1. Важен способ получения: готовые метки при обучении или поиск структуры без них.
2. Числовые расстояния по этой оси меняются относительно других; нужны согласованные масштабы.
3. Нет: трудность не определена одними размерами; нужны другие данные и задача.
4. По выбранным двум признакам явного разделения не видно; это допустимый результат, а не основание дорисовать группы.

Принимаются ответы по смыслу. Важна связь с общим знанием, а не воспроизведение деталей реального случая.

Для учителя: источники

1. Johnson, Ben-Zion, Meng, Vernon. Identifying Different Classes of Seismic Noise Signals Using Unsupervised Learning. Дата материала: 08.07.2020. Материал старше 2024 года. Подтверждающая фраза: «Temporal and spectral features describing the data are clustered».