**Модели машинного обучения подразделяются на типы в зависимости от того, как они работают**. Итак‚ три основных семейства моделей машинного обучения:

1)контролируемое обучение;

2)неконтролируемое обучение;

3) обучение с подкреплением.

**В чем разница между размеченными и неразмеченными данными.**

**Данные –** это всего лишь информация, признаки

**Признаки –** свойства или характеристики данных. В примере домашних животных признаками могут быть размер, имя, тип или вес. Однако некоторые признаки особенные, их называют **метками**.

**Метки?** Этот вариант чуть менее точен, как правило, если мы хотим спрогнозировать определенный признак на основе других, этот признак служит меткой. Например: если нам нужно узнать тип домашнего животного на основе информации о нем то метка, это тип домашнего животного (например, кошка или собака), либо если нам нужно спрогнозировать, болен питомец или здоров основываясь на симптомах и другой информации, то метка – это состояние здоровья (болен или здоров)

**Прогнозы.** Цель прогностической модели машинного обучения состоит в том, чтобы угадать метки в данных, предположение которое делает модель называется прогнозом.

**Размеченные и неразмеченные данные.** Размеченные данные – такие, которые поставляются с тегом или меткой (метка может быть как типом, так и числом). Неразмеченные данные – те, что поставляются без меток.

Изображение выглядит как зарисовка, мультфильм, млекопитающее, графическая вставка

Контент, сгенерированный ИИ, может содержать ошибки.

Размеченные и неразмеченные данные дают две ветви машинного обучения – **контролируемое и неконтролируемое**.

**Контролируемое обучение** – раздел машинного обучения, который работает с размеченными данными. Цель модели контролируемого обучения – спрогнозировать (угадать) метки.

Изображение выглядит как текст, диаграмма, Шрифт, графическая вставка

Контент, сгенерированный ИИ, может содержать ошибки.

Основа для принятия решений – это принцип воспоминания – формулирования – прогнозирования. Именно так работает контролируемое обучение. Модель сначала вспоминает набор данных о собаках и кошках. Затем формулирует модель или правило, что, по ее мнению, представляют собой собаку или кошку. Наконец, когда появляется новое изображение, модель делает прогноз о том, что, по ее мнению, является меткой изображения – собака или кошка.

Изображение выглядит как текст, зарисовка, рисунок, мультфильм

Контент, сгенерированный ИИ, может содержать ошибки.

**Числа и состояния** – это два типа данных, с которыми мы столкнемся в моделях контролируемого обучения, первый тип называется **числовыми** **данными**, второй – **категориальными.**

**Числовые данные –** это любой тип данных, в которых используются такие числа как 2,4 35 и т.д. Примеры: цены, размеры или веса.

**Категориальные данные** – это любой тип данных, в которых используются категории или состояния, такие как мужчина/женщина, кошка/собака/птица. Для этого набора данных имеется конечный набор категорий, которые нужно связать с каждой из точек данных.

**Два типа моделей контролируемого обучения**

**Регрессионные модели –** тип моделей, который прогнозирует числовые данные. Например: вес животного.

**Классификационные модели –** это тип модели, который прогнозирует категориальные данные. Ее результатом служит категория или состояние, например: тип животного – кошка или собака.

Изображение выглядит как текст, Шрифт, снимок экрана, документ

Контент, сгенерированный ИИ, может содержать ошибки.

**Неконтролируемое обучение: раздел машинного обучения, который работает с неразмеченными данными**

**Цель:** извлечь как можно больше из набора данных, который не имеет меток или целей для прогнозирования.



Даже если метки есть, мы все равно модем использовать методы неконтролируемого обучения, чтобы предварительно обработать их и более эффективно задействовать методы контролируемого обучения.

**Основные** направления **неконтролируемого обучения –** это **кластеризация, понижение размерности и генеративное обучение**.

**Алгоритм кластеризации –** алгоритмы, которые группируют данные в кластеры на основе сходств.

**Алгоритмы понижения размерности -** алгоритмы, которые упрощают данные и достоверно описывают их с меньшим кол-вом признаков.

**Генеративные алгоритмы –** алгоритмы, которые способны генерировать новые точки данных, похожие на существующие.

Изображение выглядит как текст, Шрифт, снимок экрана, черно-белый

Контент, сгенерированный ИИ, может содержать ошибки.

**Наиболее важные алгоритмы кластеризации (ИЗУЧИТЬ САМОСТОЯТЕЛЬНО)**

**Изображение выглядит как текст, Шрифт, снимок экрана, число

Контент, сгенерированный ИИ, может содержать ошибки.**

**Алгоритмы понижения размерности:**

1)Нахождение схожих признаков и их удаление

Пример: Пусть точка данных – это строка, а признаки – столбцы, тогда при нахождении одинаковых признаков, мы удаляем не нужные столбцы

2)Матричная факторизация и разложение по сингулярным значениям:

Предположим, что у нас есть большая таблица данных, где точка данных – строка, а признаки – столбцы, следовательно мы можем использовать кластеризацию для уменьшения кол-ва строк и понижения размерности для уменьшения кол-ва столбцов. Но есть способ, который делает это все одновременно, это – **РАЗЛОЖЕНИЕ МАТРИЦ НА МНОЖИТЕЛИ** и **РАЗЛОЖЕНИЕ ПО СИНГУЛЯРНЫМ ЗНАЧЕНИЯМ.** Эти два алгоритма преобразуют большую матрицу данных в произведение меньших матриц. **(ТАКЖЕ ПРОЙТИ САМОСТОЯТЕЛЬНО)**

**Генеративное машинное обучение**

**Генеративное машинное обучение** (генерирование реалистичных лиц, изображений и т.д)

**Основной принцип:** модель, учитывая набор данных, может вводить новые точки данных, выглядящие как образцы из исходного набора.

**Обучение с подкреплением**

**Обучение с подкреплением –** еще один тип машинного обучения, при котором данные изначально не передаются, и мы должны заставить компьютер выполнить задачу. Вместо данных модель получает среду и агента, который должен в ней перемещаться. У агента есть цель или набор целей.

