Вопросы к первому уроку

1. Назовите основные проблемы машинного обучения?

Недостаточный размер обучающих данных (недостаточная история наблюдений, возможно низкое их качество -> недостаточный размер качественных данных)

Нерепрезентативная выборка (понял так, будто выборка данных не описывает сути исследуемого объекта)

Зашумленность, пропуски в данных. (например, низкое качество измерений. Будет сложно оценить потери мощности и энергии в линии передачи, если точность измерения имеет одинаковый порядок с передаваемой мощностью, а не с величиной потерь, или же данные не синхронизированы по времени)

Несущественные признаки (ошибочное выделение признаков. В действительности функция не является функцией данного аргумента или является ею «слабо»)

Переобучение (алгоритм запомнил определенные признаки подобно расположению правильных ответов в тесте, а не содержание правильных ответов)

Недообучение (…)

1. В чем разница между параметрами модели и гиперпараметрами?

Параметр – внутренняя характеристика модели. Изменяющаяся в процессе обучения. Гиперпараметр – внешняя характеристика. По всей видимости задаётся в самом начале и определяет качество модели (или скорость ее обучения). Если сравнить с итерационным способом решения уравнения, то гиперпараметр – начальное приближение, параметр – полученное значение.

1. Что такое помеченный обучающий набор? В каких моделях используется такой набор?

Это набор данных, выделенный для обучения и модели и имеющий пометки (например датасет с 10 000 комментариев с сайта Пикабу, а напротив каждого стоит пометка токсичный/ не токсичный) Применяется в моделях обучения с учителем.

1. Что такие динамическое обучение?

Обучение модели непрерывным потоком данных. Как пример модель может делать прогноз на сутки вперед используя наблюдения за год «назад». Потом модель видит фактические данные по прошествии суток. Наверное, она может их сравнить со своим собственным прогнозом, скорректировать параметры, а новые данные положить в архив истории наблюдений

1. Как вы думаете задача кредитного скоринга эта задача обучения с учителем или без учителя?

Сказал бы, что с учителем. Это классификационная задача.

1. Как вы думаете задача определения спама эта задача обучения с учителем или без?

Можно и без учителя. Небольшой процент пропущенных писем может не стоить того, чтобы предварительно промаркировать датасет. Но при наличии датасета из пункта 3, можно обучать и с учителем))

1. Какой алгоритм можно применить что бы разделить данные на несколько групп, которые обладают схожими свойствами?

Кластеризация. Используется при наличие непомеченных данных. Алгоритм замечает принадлежность данных к одной или нескольким группам и соответствующим образом группирует.

1. Как вы понимаете определение «машинное обучение». Напишите своими словами.

Думаю, это можно сравнить с системой с положительной обратной связью. В электроэнергетике, например, схожими свойствами обладает батарея статических конденсаторов. Их применяют для генерации реактивной мощности и повышения напряжения. Положительная обратная связь здесь в том, что бОльшее приложенное напряжение вызывает генерацию бОльшей мощности -> генарция бОльшей мощности вызывает бОльшее повышение напряжения. В итоге данный процесс приходит в равновесие. В машинном обучении, наверное, этот процесс будет происходить до тех пор, пока модель не научится извлекать из подводимой к ней информации максимум выгоды (хотя само значение максимума, наверное, вовсе не достижимо. К тому же модель может вовсе не обучаться, не приобретать опыт. Такой случай можно сравнить с понижением напряжения)

1. Какой алгоритм обучения вы бы использовали если бы хотели научить ИИ играть в Доту2\ориентироваться робота на неизвестной территории?

Обучение с подкреплением. ИИ штрафуется за плохие нежелательные действия (стояние на месте, гибель, ловушки, промахи и прочее), вознаграждается за верные действия (нанесенный урон, полученные очки и прочее)

1. Каковы две наиболее распространенные задачи обучения с учителем?

Задача классификации (спам писем), задача прогнозирования (прогнозирование целевого ценового значения)

1. Какое разделение профессий в области DS вы знаете?

Инженер - много кодит

Аналитик – кодит и анализирует

Специалист – много анализирует

1. Что такое переменные окружения?

Набор пар типа «Переменная» – «Значение». Хранят пути доступа к исполняемым файлам. Соответственно переменная название исполняемого файла, значение путь к нему.