Вопросы к первому уроку

1. Назовите основные проблемы машинного обучения?

Недостаточный размер обучающих данных

Нерепрезентативные обучающие данные

Зашумленность / пропуски в данных

Несущественные признаки (выбор признаков / создание новых признаков)

Переобучение

Недообучение

1. В чем разница между параметрами модели и гиперпараметрами?

**Параметры** модели – это параметры, вычисляемые в процессе обучения модели.

**Гиперпараметры** модели — параметры, значения которых задаются до начала обучения модели и не изменяются в процессе обучения. Это внешние по отношению к модели параметры, значения которых невозможно оценить по данным. Гиперпараметры используются для управления процессом обучения.

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

Помеченный обучающий набор – это примеры желательных решений для каждого образца.

Применяется в моделях обучения с учителем.

1. Что такое динамическое обучение?

Динамическое обучение – это постепенное обучение в процессе работы модели на непрерывном и изменяющемся потоке данных либо при обработке гигантских данных.

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

Я думаю, что это задача обучения **без учителя**, здесь нет примера достоверного правильного ответа, на выходе мы получаем вероятность риска для того, кто выдаёт кредит.

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

Я думаю, что это задача обучения **с учителем**, так как есть примеры правильных ответов.

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

Нужно выполнить кластеризацию, так как количество групп неизвестно («несколько»).

Для этого нужно использовать обучение без учителя.

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

Машинное обучение – это наука о том, как программировать компьютерные системы, чтобы они могли обучаться при обработке данных и корректировать свои действия без вмешательства человека.

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

Обучение с подкреплением.

1. Каковы две наиболее распространенные задачи обучения с учителем?

Классификация и прогнозирование

1. Какое разделение профессий в области DS вы знаете?

data-аналитик, data-инженер, data-scientist

1. Что такое переменные окружения?

Переменные окружения – это переменные, в которых указаны настройки внешних условий и обстоятельств, в которых будет работать модель (пути к базам данных, тип операционной системы, версии ПО)