ВОПРОСЫ К ПЕРВОМУ УРОКУ

1. **Назовите основные проблемы машинного обучения?**

Основными проблемами машинного обучения являются:

- недостаточный размер обучающих данных;  
- нерепрезентативные обучающие данные;  
- данные плохого качества(;зашумленность/ пропуски данных);  
- несущественные признаки;  
- переобучение;  
- недооучение;

1. **В чем разница между параметрами модели и гиперпараметрами?**

Значение параметров модели можно оценить на основе данных, а значение гиперпараметра нельзя по данным оценить.

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

Помеченный обучающий набор используется для обучения системы МО с учителем. В таком случае в обучающем наборе данных, которые включаются желательные решения, которые называются метками.

Помеченный обучающий набор используется в следующих моделях:

- линейная регрессия;

- логистическая регрессия;

- метод опорных векторов;

- дерево приятия решений и случайные леса;

- k ближайших соседей;

- нейронные сети.

1. **Что такие динамическое обучение?**

Динамическое обучение это вид машинного обучения, в котором системе МО данные предоставляются последовательно предоставляются небольшие наборы данных(мини-пакеты).

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

Обучение с учителем, чтобы были определены признаки по которым можно было бы сделать вывод является ли человек(который хочет взять кредит) платежеспособным или нет.

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

Обучение с учителем. Чтобы в предоставленном наборе данных были определены признаки, по которым определяла, является ли письмо спамом или нет(Пример: наличие слов «выигрыш», «призы» и тд)

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

Алгоритм кластеризации.

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

Машинное обучение это способ«обучения» программы на основе загруженных данных. Т.е. в программу загружается датасет и система на основе выбранной модели(алгоритма) уже без дополнительного участия со стороны пользователя делает прогнозы/выдает некий результат.

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?**

Распространенными задачами обучения с учителем являются:

- классификация;

- прогнозирование целевого числового значения.

1. **Какое разделение профессий в области DS вы знаете?**

- Data Analyst;

- Data Engineer;

- Data Scientist;

1. **Что такое переменные окружения?**

Переменная окружения — это текстовая переменная операционной системы, хранящая какую-либо информацию(пример: данные о настройках системы или сведения о текущем пользователе).