Ответы на вопросы к первому уроку

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

«Плохие» данные:

* Недостаточный размер обучающих данных
* Нерепрезентативные обучающие данные
* Данные плохого качества
* Несущественные признаки

«Плохие» алгоритмы:

* Переобучение обучающими данными
* Недообучение обучающими данными

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

**Параметр модели** - это переменная конфигурации, которая является внутренней по отношению к модели и значение которой можно оценить на основе данных.

Параметры являются частью модели, которая извлекается из исторических данных обучения.

**Гиперпараметр модели** - это конфигурация, внешняя по отношению к модели, значение которой невозможно оценить по данным.

Хорошее практическое правило для преодоления этой путаницы заключается в следующем:

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

При обучении с учителем обучающие данные, поставляемые алгоритму, включают для каждого образца желательные решения, называются метками или помеченным обучающим набором).

1. **Что такие динамическое обучение?**

При **динамическом обучении** вы обучаете систему постепенно за счет последовательного предоставления ей образцов данных либо по отдельности, либо небольшими группами, называемыми мини-пакетами. Каждый шаг обучения является быстрым и недорогим, так что система может узнавать о новых данных на лету по мере их поступления. Динамическое обучение отлично подходит для систем, которые получают данные в виде непрерывного потока (скажем, курс акций) и должны адаптироваться к изменениям быстро или самостоятельно. Вдобавок динамическое обучение будет хорошим вариантом в случае ограниченных вычислительных ресурсов: после того, как система динамического обучения узнала о новых образцах данных, она в них больше не нуждается, а потому может их отбросить (если только вы не хотите иметь возможность производить откат к предыдущему состоянию и «воспроизведение» данных).

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

Оценивание кредитоспособности заёмщиков – это задача обучения с учителем, так как это задача на классификацию. Обучающая выборка составляется из заёмщиков с известной кредитной историей. В простейшем случае принятие решений сводится к классификации заёмщиков на два класса: «хороших» и «плохих». Кредиты выдаются только заёмщикам первого класса.

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

Определения спама – это задача обучения с учителем, так как это задача на классификацию.

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

Задача [кластеризации](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F) (clustering) заключается в том, чтобы сгруппировать объекты в [кластеры](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%BB%D0%B0%D1%81%D1%82%D0%B5%D1%80&action=edit), используя данные о попарном сходстве объектов, и решается обучением без учителя  (unsupervised learning), если неизвестны параметры группировки. В случае, если они известны, задача разделения данных на несколько групп решается является классификацией и решается обучением с учителем.

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

Машинное обучение – это построение систем, способных учится на основе данных. Обучение предполагает за собой улучшение процесса и его эффективности.

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

**Обучение с подкреплением** ([англ.](https://ru.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9_%D1%8F%D0%B7%D1%8B%D0%BA) reinforcement learning) — один из способов [машинного обучения](https://ru.wikipedia.org/wiki/%D0%9C%D0%B0%D1%88%D0%B8%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BE%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5), в ходе которого испытуемая система (агент) обучается, взаимодействуя с некоторой средой. С точки зрения [кибернетики](https://ru.wikipedia.org/wiki/%D0%9A%D0%B8%D0%B1%D0%B5%D1%80%D0%BD%D0%B5%D1%82%D0%B8%D0%BA%D0%B0), является одним из видов [кибернетического эксперимента](https://ru.wikipedia.org/wiki/%D0%9A%D0%B8%D0%B1%D0%B5%D1%80%D0%BD%D0%B5%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B9_%D1%8D%D0%BA%D1%81%D0%BF%D0%B5%D1%80%D0%B8%D0%BC%D0%B5%D0%BD%D1%82). Откликом среды (а не специальной системы управления подкреплением, как это происходит в [обучении с учителем](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D1%81_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D0%B5%D0%BC)) на принятые решения являются сигналы подкрепления, поэтому такое обучение является частным случаем [обучения с учителем](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D1%81_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D0%B5%D0%BC), но учителем является среда или её модель. Также нужно иметь в виду, что некоторые правила подкрепления базируются на неявных учителях, например, в случае [искусственной нейронной среды](https://ru.wikipedia.org/wiki/%D0%98%D1%81%D0%BA%D1%83%D1%81%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%B0%D1%8F_%D0%BD%D0%B5%D0%B9%D1%80%D0%BE%D0%BD%D0%BD%D0%B0%D1%8F_%D1%81%D0%B5%D1%82%D1%8C), на одновременной активности формальных нейронов, из-за чего их можно отнести к [обучению без учителя](https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%83%D1%87%D0%B5%D0%BD%D0%B8%D0%B5_%D0%B1%D0%B5%D0%B7_%D1%83%D1%87%D0%B8%D1%82%D0%B5%D0%BB%D1%8F).

1. **Каковы две наиболее распространенные задачи обучения с учителем?**

* Задача [классификации](http://www.machinelearning.ru/wiki/index.php?title=%D0%9A%D0%BB%D0%B0%D1%81%D1%81%D0%B8%D1%84%D0%B8%D0%BA%D0%B0%D1%86%D0%B8%D1%8F) (classification) отличается тем, что множество допустимых ответов конечно. Их называют метками классов (class label). Класс — это множество всех объектов с данным значением метки.
* Задача [регрессии](http://www.machinelearning.ru/wiki/index.php?title=%D0%A0%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D1%8F) (regression) отличается тем, что допустимым ответом является действительное число или числовой вектор.

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

* Аналитик данных
* Дата-инженер
* Архитектор баз данных
* ML/DL-исследователь
* ML/DL-инженер
* Data-программист

1. **Что такое переменные окружения?**

Переменные окружения - это именованные переменные, содержащие текстовую информацию, которую могут использовать запускаемые программы. С помощью переменных окружения можно получать различные параметры приложение и секретные ключи.