**Способы машинного обучения**

(1 слайд)

(2 слайд)

**BPC:** В нашем выступлении речь пойдет о различных способах машинного обучения. В современном мире мы все больше и больше используем компьютеры для ускорения процесса решения разных задач. Например, по утрам смотрим прогноз погоды или, садясь в свою машину, говорим Google помощнику проложить маршрут до какого-то нужного нам места. Но вы когда-нибудь задумывались о том, как происходит предсказание погоды или распознавание телефоном голоса человека? Никто не будет в своей программе записывать все возможные варианты движения облаков, или различные звуки человеческого голоса, изменения давления воздуха или его температуры. Это сделать невозможно. Конечно, во всем этом присутствует доля вычисления по точно заданным формулам и действия по определённым алгоритмам, но во многих случаях используется машинное обучение, которое автоматизирует и очень сильно ускоряет данный процесс.

**PS:** Машинное обучение – это быстро развивающаяся технология, которую сейчас внедряют в огромное количество систем и используют повседневно. Поэтому мы и решили изучить эту сферу компьютерных алгоритмов.

(3 слайд)

**BPC:** Давайте разберемся поподробнее, что такое машинное обучение и где его применяют. Машинное обучение – это различные алгоритмы для анализа данных, предназначенные для тренировки аналитической системы в процессе решения схожих задач, чтобы потом она могла отвечать на подобные запросы пользователей. Сфера применения машинного обучения очень велика. Мы уже приводили пример с распознаванием речи. Также оно может использоваться для обнаружения нежелательных писем и спама в почте или различной диагностики. И это ещё далеко не все возможности!

**PS:** Первым делом расскажем о том, как вообще проходит машинное обучение. Сначала аналитической системе на вход подаются тренировочные данные – множество схожих задач. Каждая из них проходит через определенный алгоритм. Получившийся на выходе ответ для каждой задачи сравнивается с тем, который должен быть, и, в зависимости от результата, этот алгоритм отлаживается и исправляется для того, чтобы правильный ответ давался в наибольшем количестве случаев. Затем аналитическая система получает тестовые данные, на которых оценивается ее точность. После этого, в зависимости от полученного результата точности, аналитическую систему улучшают и исправляют недочеты.

(4 слайд)

**BPC:** Существует очень много алгоритмов для машинного обучения. И все они по-своему подходят для решения каких-то задач. Но каждый из них применяют не везде и используют только в некоторых местах, для решения определенных проблем. Мы решили разобраться в устройстве и применении алгоритмов на примере четырех: логистической регрессии, метода опорных векторов, нейронной сети и дерева принятия решений. В нашем исследовании мы пытались использовать данные алгоритмы для решения схожих задач, а потом на тестовых данных сравнивали их точность. Затем, на основе полученных результатов нашего эксперимента анализировали, где и для чего может применяться каждый из них. Но сначала расскажем поподробнее о самих алгоритмах.

(5 слайд)

**BPC:** Алгоритм логистической регрессии заключается в преобразовании зависимости тренировочных данных в логистическую функцию, которая преобразует все значения в значения от нуля до единицы. То есть, при обучении аналитическая система просто подбирает такие коэффициенты для данной функции, чтобы на выходе получалась наибольшая точность. Этот алгоритм хорошо подходит для задач, в которых нужно выяснить принадлежность какого-то объекта к данному классу, а точки данных можно однозначно разделить кривой на 2 типа.

(6 слайд)

**PS:** Метод опорных векторов – это один из наиболее популярных алгоритмов машинного обучения. В нем, как и в логистической регрессии, тоже подбираются значения коэффициентов уравнения кривой. Но здесь эта кривая является гиперплоскостью, разделяющей пространство входных значений. Во время обучения аналитической сети ее коэффициенты подбираются так, чтобы получившаяся функция наилучшим образом разделяла точки данных по двум классам. Идеальной (оптимальной) гиперплоскостью называется такая, от которой расстояние до ближайших точек будет максимальным. Только эти точки играют роль при обучении аналитической сети. Они называются опорными векторами.

(7 слайд)

**BPC:** Нейронная сеть – самый известный и очень распространенный метод машинного обучения. Она содержит в себе три типа слоев: входной, скрытый и выходной. В каждом из них находятся нейроны, соединенные синапсами, имеющими свой вес, на который умножается поступившее в них число. Данные аналитическая система получает через входной слой. Затем полученные результаты отправляются в скрытый слой. В нем как раз и происходит весь вычислительный процесс. При прохождении через синапсы, каждое из полученных значений умножается на какое-то число - вес данного синапса. Из выходного слоя мы получаем ответ на задачу. Процесс обучения заключается в оптимальном подборе весов синапсов, которые сначала генерируются произвольно, а затем, в процессе тренировки, уже начинают корректироваться в зависимости от различия между правильным и текущим ответами. Этот алгоритм имеет достаточно большую сферу применения, так как можно изменять множество разных параметров.

(8 слайд)

**PS:** Одним из модификаций прошлого алгоритма являются сверточные нейронные сети. Обычно они применяются для распознавания объектов на картинках. В них, помимо нейронов и синапсов, используются так называемые сверточные слои, в которых выделяются схожие признаки на данном изображении. Это делается тоже путем умножения чисел, в данном случае – значения пикселей, на веса ядер свертки.

(9 слайд)

**BPC:** Ещё один алгоритм машинного обучения - дерево принятия решений. Каждый его узел представляет собой какую-то входную переменную и точку разделения для нее. В примере на презентации таковыми являются требования для выбора места работы. Также существуют листовые узлы, которые применяются для предсказаний путем прохода от корня дерева до данного узла и вывода значения на нем. В примере на презентации таковыми являются положительные и отрицательные ответы на предложение о вакансии. Дерево принятия решений – это быстро обучающийся алгоритм, который может применяться для решения большого круга задач. Кроме того, он не требует особой подготовки данных, что существенно облегчает процесс его тренировки.

(10 слайд)

**PS:** Для проведения исследования мы использовали язык программирования Python 3, потому что для него существует множество инструментов для простой работы с машинным обучением. Мы выбрали фреймворки scikit-learn и Keras. Первый позволяет использовать множество алгоритмов работы с большими данными, а второй даёт возможность работы со сверточным нейронными сетями. Мы использовали среду разработки PyCharm, которая является одной из самых удобных для языка Python.

(11 слайд)

**PS:** А теперь расскажем о результатах нашего исследования. После применения данных способов машинного обучения на MNIST (базе данных рукописных цифр) выяснилось, что самыми точными алгоритмами являются метод опорных векторов и сверточная нейронная сеть, в меньшей степени – логистическая регрессия и дерево принятия решений. Но, исходя из результатов видно, что у всех алгоритмов достаточно высокая точность, более 95 процентов. Так же мы применили их для распознавания объектов на изображениях на примере кошек и собак. Результат показал, что в данной сфере сверточная нейронная сеть показывает более высокий результат точности, а логистическая регрессия – самый низкий.

(12 слайд)

**BPC:** На основе исследования можно сделать вывод, что для распознавания объектов на изображениях лучше всего подходят сверточные нейронные сети, потому что они находят признаки в различных частях изображения. Но если вам нужен неплохой показатель точности при достаточно быстром обучении на простых задачах, то лучше всего использовать метод опорных векторов или дерево принятия решений.

**PS:** Нельзя однозначно сказать, что какой-то алгоритм лучше других. Каждый из них создан для решения какой-то определенной задачи и хорош в данной области по-своему. Так что перед применением какого-то алгоритма лучше всего сначала подумать о его совместимости с данной задачей.

(13 слайд)

(14 слайд)