

Машинное обучение

ТЕЗАУРУС

Что же такое машинное обучение?

Машинное обучение считается ветвью искусственного интеллекта, основная идея которого заключается в том, чтобы компьютер не просто использовал заранее написанный алгоритм, а сам обучился решению поставленной задачи.

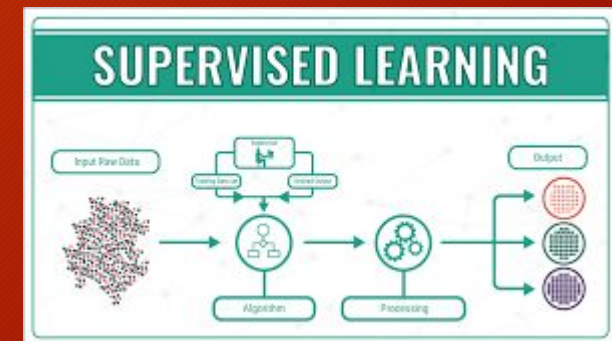


Обучение с учителем

Supervised learning

Один из разделов машинного обучения, посвященный решению следующей задачи. Имеется множество объектов (ситуаций) и множество возможных ответов (откликов, реакций). Существует некоторая зависимость между ответами и объектами, но она неизвестна. Известна только конечная совокупность прецедентов — пар «объект, ответ», называемая обучающей выборкой. На основе этих данных требуется восстановить зависимость, то есть построить алгоритм, способный для любого объекта выдать достаточно точный ответ. Для измерения точности ответов определённым образом вводится функционал качества.

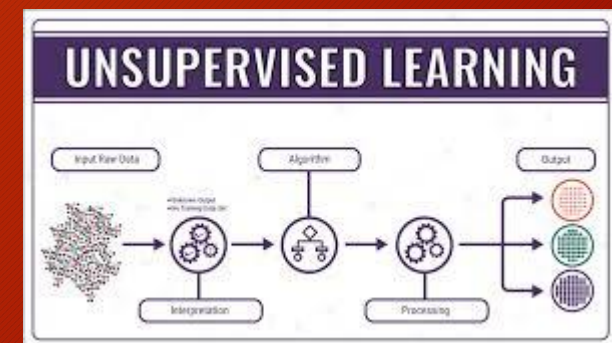
Под учителем понимается либо сама обучающая выборка, либо тот, кто указал на заданных объектах правильные ответы.



Обучение без учителя

Unsupervised learning

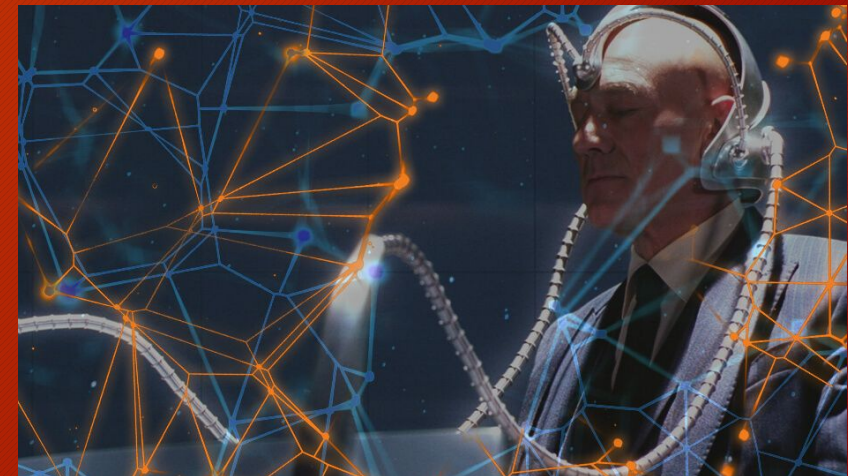
Один из разделов машинного обучения . Изучает широкий класс задач обработки данных, в которых известны только описания множества объектов (обучающей выборки), и требуется обнаружить внутренние взаимосвязи, зависимости, закономерности, существующие между объектами.



Нейронная сеть

Это последовательность нейронов, соединенных между собой синапсами. Структура нейронной сети пришла в мир программирования прямиком из биологии. Благодаря такой структуре, машина обретает способность анализировать и даже запоминать различную информацию. Нейронные сети также способны не только анализировать входящую информацию, но и воспроизводить ее из своей памяти. Другими словами, нейросеть это машинная интерпретация мозга человека, в котором находятся миллионы нейронов передающих информацию в виде электрических импульсов.

Фактически же, нейрон в искусственной нейронной сети представляет собой математическую функцию которой на вход приходит какое-то значение и на выходе получается значение, полученное с помощью той самой математической функции.

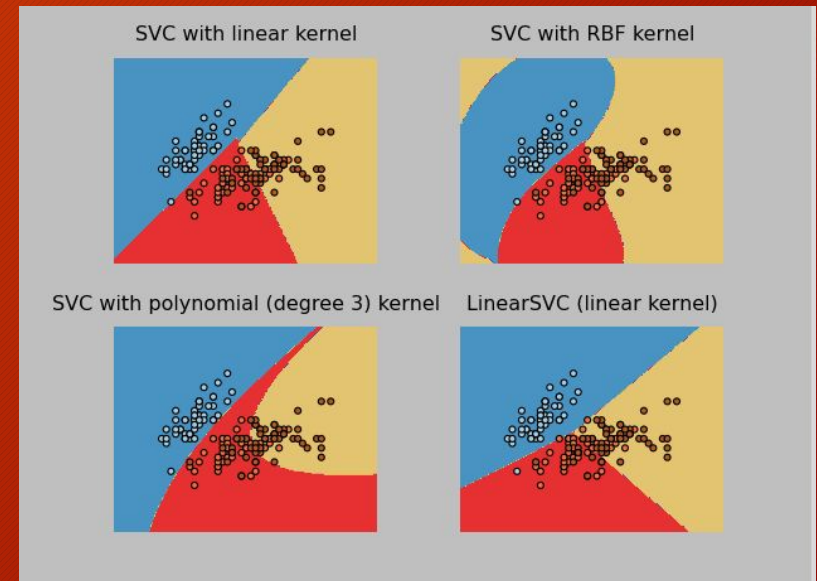


Метод опорных векторов

Особым свойством метода опорных векторов является непрерывное уменьшение эмпирической ошибки классификации и увеличение зазора.

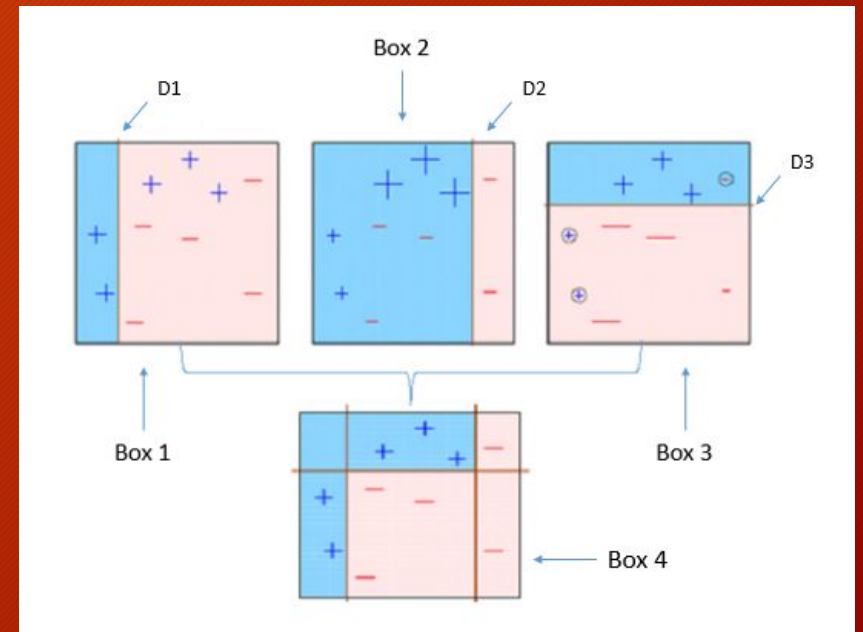
Основная идея метода опорных векторов — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей наши классы.

Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора.



БУСТИНГ

Процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов. Бустинг представляет собой жадный алгоритм построения композиции алгоритмов. В течение последних 10 лет бустинг остаётся одним из наиболее популярных методов машинного обучения, наряду с нейронными сетями и машинами опорных векторов. Основные причины — простота, универсальность, гибкость (возможность построения различных модификаций), и, главное, высокая обобщающая способность.



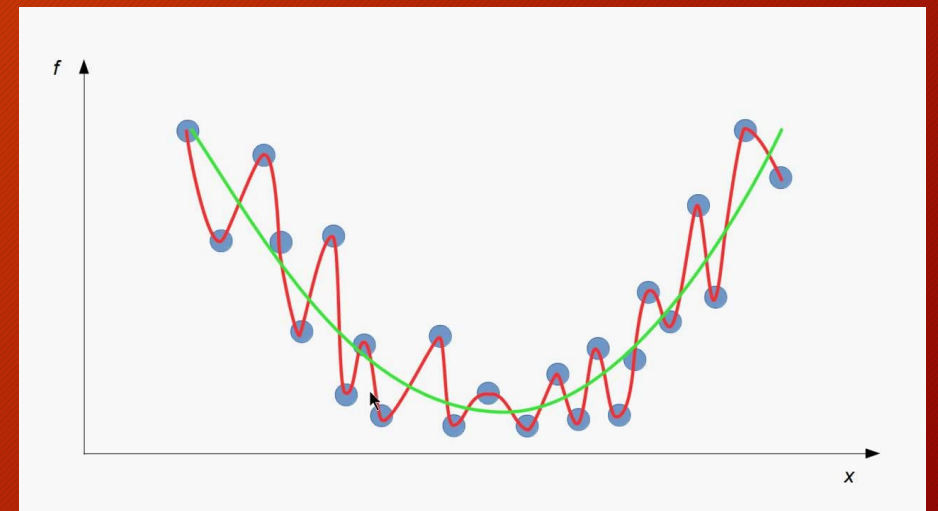
Задачи машинного обучения

- Регрессии
- Классификации
- Кластеризации
- Уменьшения размерности
- Выявления аномалий



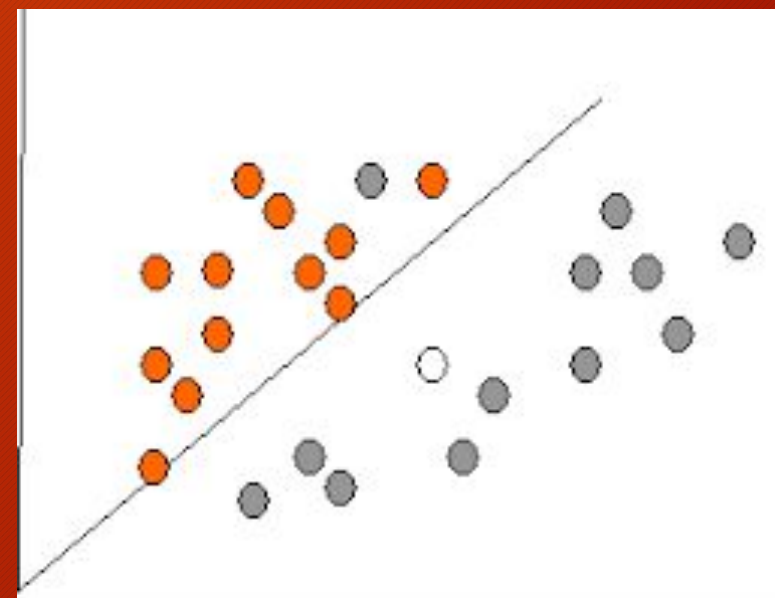
Задачи регрессии

Метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной). Регрессионная модель есть функция независимой переменной и параметров с добавленной случайной переменной. Параметры модели настраиваются таким образом, что модель наилучшим образом приближает данные.



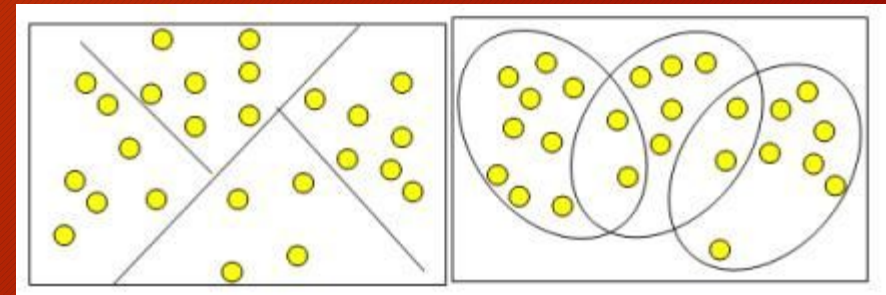
Задача классификации

Один из разделов машинного обучения, посвященный решению следующей задачи. Имеется множество объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется обучающей выборкой. Классовая принадлежность остальных объектов не известна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.



Задача кластеризации

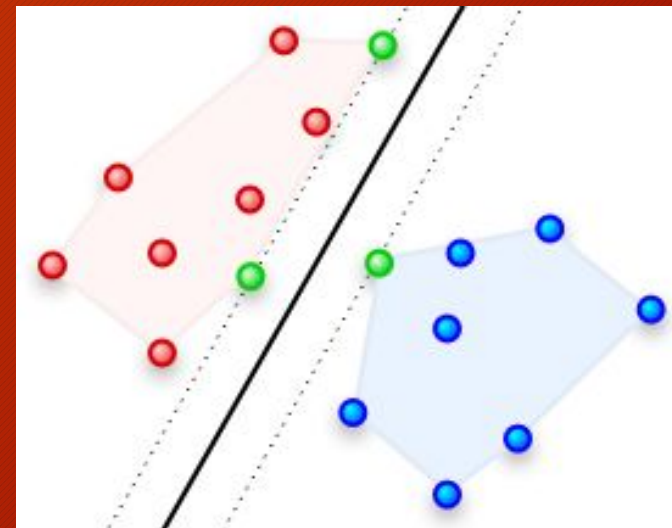
Задача разбиения заданной выборки объектов (ситуаций) на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.



Задача уменьшений размерности

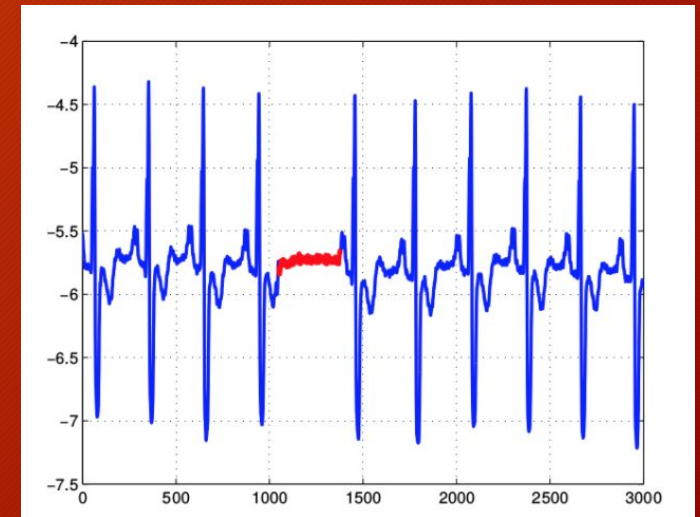
Это один из основных алгоритмов машинного обучения. Позволяет уменьшить размерность данных, потеряв наименьшее количество информации. Применяется во многих областях, таких как распознавание объектов, компьютерное зрение, сжатие данных и т. п.

Вычисление главных компонент сводится к вычислению собственных векторов и собственных значений ковариационной матрицы исходных данных или к сингулярному разложению матрицы данных.



Задача выявления аномалий

В анализе данных есть два направления, которые занимаются поиском аномалий: детектирование выбросов и «новизны». Как и выброс «новый объект» — это объект, который отличается по своим свойствам от объектов (обучающей) выборки. Но в отличие от выброса, его в самой выборке пока нет (он появится через некоторое время, и задача как раз и заключается в том, чтобы обнаружить его при появлении). Например, если вы анализируете замеры температуры и отбрасываете аномально большие или маленькие, то вы боретесь с выбросами. А если вы создаёте алгоритм, который для каждого нового замера оценивает, насколько он похож на прошлые, и выбрасывает аномальные — вы «боретесь с новизной».



Спасибо за внимание!!!!

