Линейный классификатор

Задача

|  |  |
| --- | --- |
| **X** | **y** |
| -2 | -1 |
| -1 | 1 |
| 0 | -1 |
| 1 | 1 |
| 2 | 1 |

Данный вид классификаторов основан на определении знака функции, вида:

где – признаки, описывающие объект, вектор ( – веса соответствующих признаков, – порог принятия решения, n – количество признаков.

При решении задачи линейной классификации определим два класса с метками 1 и -1 соответственно. Тогда решением задачи классификации в данном случае будет являться следующее выражения:

Рассмотрим пример. Предположим, имеются данные по объектам, описывающимся двумя признаками и результирующим признаком, принимающим значения 1 или -1.

Построим график этих данных.

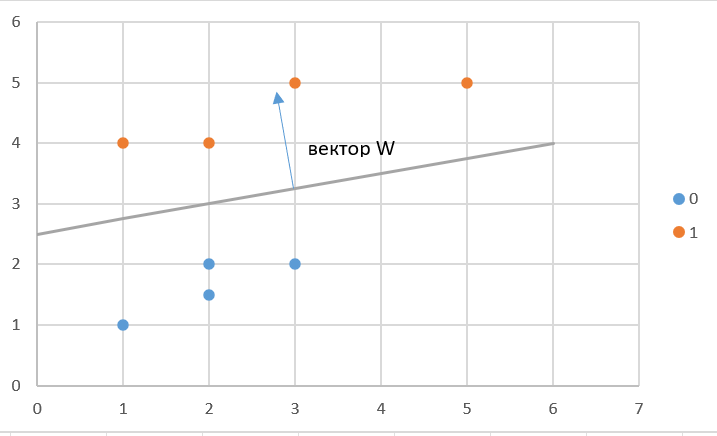
Как видно на графике, можно попытаться построить такую прямую, которая сможет разделить данные на классы, то есть ниже прямой будут лежать объекты, имеющие метку класса -1, а выше +1. Нарисуем теоретическую прямую.

Как видно из рисунка, прямая удачно разделила наши данные на 2 класса. Все объекты, лежащие ниже прямой, будут относиться к классу -1. Тогда решение задачи линейного фильтра будет выглядеть следующим образом:

Если , то объект относится к классу +1,

Если , то объект относится к классу -1.

Таким образом, если подставить конкретные значения x и y в данное уравнение, то знак полученного выражения будет определять принадлежность к классу. На самом деле, является скалярным произведением вектора X и вектора W. Построенная прямая определяется выражением Поэтому вектор W является ортогональным по отношению к построенной прямой В связи с этим, знак скалярного произведения будет определяться косинусом угла между ними. Если объект находится ниже прямой, то угол будет тупым, то есть значение косинуса между рассматриваемыми векторами будет иметь знак -.

**

Как определить оптимальное положение разделяющей прямой? Для этого необходимо определить функцию потерь.

Функция ошибается в 2-х случаях:

1. , но объект из класса -1;
2. , но объект из класса +1.

Тогда функция потерь M называется отступом и будет иметь следующий вид:

Или:

Но данная функция не является непрерывной, поэтому мажорирующую функцию:

Мажорирующая функция F — это функция, которая используется для оценки или "мажорирования" другой функции, обеспечивая верхнюю границу для её значений, то есть:

Такой функцией может быть F = (1-M)2

Рассчитаем отступы:

|  |  |
| --- | --- |
| **M** | **Отступ** |
| M1 |  |
| M2 |  |
| M3 |  |
| M4 |  |
| M5 |  |

Тогда нужно минимизировать выражение:

… =

=

Оптимизируем данную функцию, взяв частные производные по и и приравняем их к 0:

Решением уравнения будут два значения:

*Следовательно, классификатор будет таким: если 0.4x+0.2>0, то объект из класса 1 (иначе из класса -1).*

Разобранный ранее пример очень простой, и найти точку минимума (оптимальные значения весов wi ) несложно. Однако если переменных много, то система разрастается и тяжело найти решение такой системы.

Поэтому будем использовать другой метод оптимизации.

В качестве метода оптимизации возьмем метод стохастического градиентного спуска. Метод стохастического градиентного спуска предполагает следующий алгоритм обновления весов:

где – скорость обучения модели (шаг обучения), обычно маленькое число, например 0.05

Теперь решим наш пример с использованием градиентного спуска:

Градиентом по определению будет частная производная по нашим параметрам и , а их мы уже нашли:

Возьмем произвольную точку. Например, (1,1), то есть:

Считаем значение градиента в этой точке, подставляя вместо и единицы. А далее считаем новые координаты точки:

Новая точка будет

Рассчитаем еще один шаг градиента

И таким образом выполняем шаги градиентного спуска до тех пор, пока коэффициенты не перестанут меняться.