**ADALINE, правило Хэбба**

Игнатенко Д. 401-И

Пусть дана обучающая выборка: множество входных значений **** и множество выходящих значений . Необходимо по этим данным построить ADALINE, которая допускает наименьшее количество ошибок на этой обучающей выборке. Обучение ADALINE заключается в подборе наилучших значений вектора весов **,** при котором достигается минимум аппроксимированного эмпирического риска:



где  – скалярное произведение, – функция потерь.

В данном случае используется функция потерь следующего вида:



Для подбора вектора весов применим метод **стохастического градиента**. Но для начала проведем нормализацию признаков, чтобы итерационный процесс не оказался «парализованным»:



Сначала необходимо инициализировать веса небольшими случайными значениями , где – размерность пр-ва.

Далее нужно вычислить значение функционала .

Выбрать случайный объект , вычислить ошибку , сделать шаг градиентного спуска , где  – темп обучения.

Затем оценить новое значение функционала , где – параметр сглаживания. Процесс продолжать, пока  не стабилизируется.

Для оптимизации алгоритма обучения можно использовать правило Хэбба, которое заключается в том, чтобы обновлять вектор весов только в том случае, когда выбранный объект имеет отрицательный отступ.

Результат работы программы приведен на рисунках ниже.

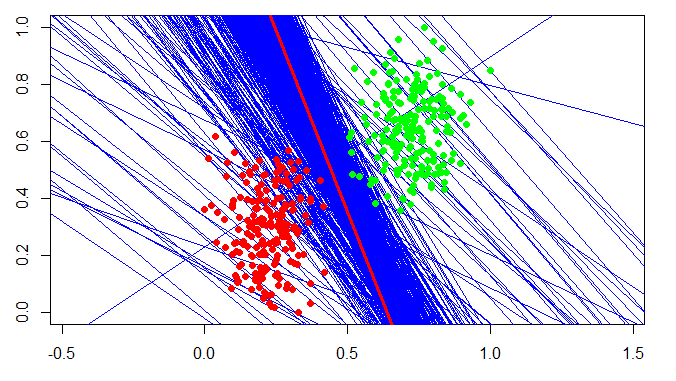


Рисунок 1. ADALINE – 867 изменений вектора весов

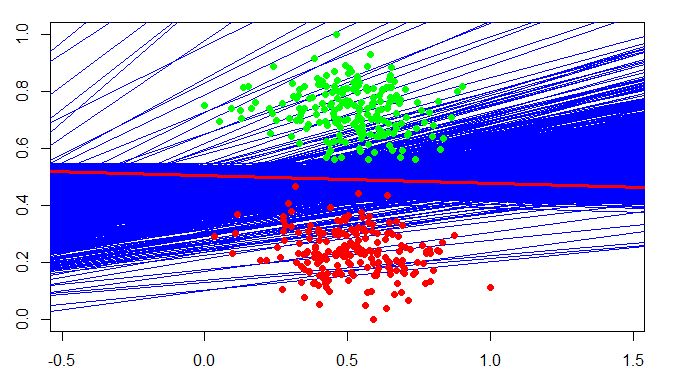


Рисунок 2. ADALINE – 779 изменений вектора весов

**Правило Хэбба**

Если ответ алгоритма совпадает с меткой класса, то вектор весов обновлять не нужно. Иначе, если объект имеет отрицательный отступ , то обновление в методе стохастического градиента происходит по следующему правилу:

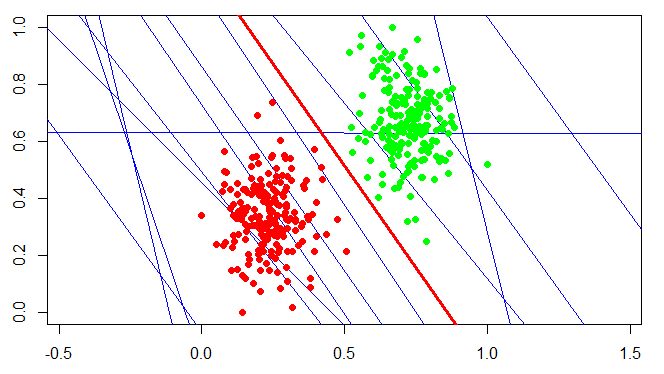
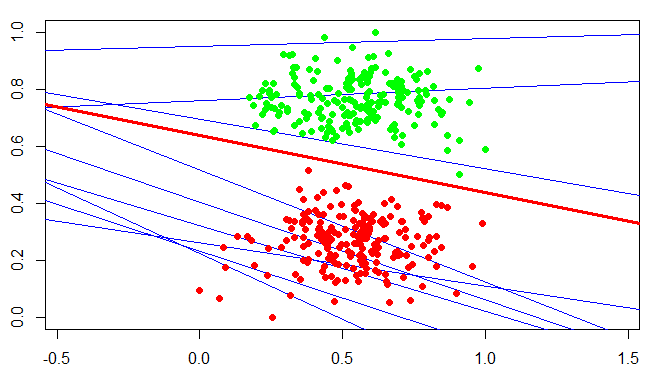


Рисунок 4. Правило Хэбба – 13 изменений вектора весов

Рисунок 3. Правило Хэбба – 17 изменений вектора весов

**Вывод**:

Как видно из рисунков, если обновлять вектор весов только на ошибочных объектах, то подбор оптимальной разделяющей прямой происходит за меньшее число шагов, но при этом без потери качества классификации.

**Преимущества** метода стохастического градиента:

* Простая реализация.
* Подходит для динамического обучения, когда обучающие объекты поступают потоком, и вектор весов обновляется при появлении каждого объекта.
* Метод позволяет настраивать веса на избыточно больших выборках, за счёт того, что случайной подвыборки может оказаться достаточно для обучения.

**Недостатки**:

* Функционал, как правило, многоэкстремальный, и процесс может сходиться к локальному минимуму, сходиться очень медленно или не сходиться вовсе.
* При большой размерности пространства n или малой длине выборки ℓ возможно переобучение.
* Если функция потерь имеет горизонтальные асимптоты, то процесс может попасть в состояние «паралича» (решается нормализацией признаков).