



Линейные методы

Евгений Борисов

Линейные методы

методы ML

- *метрические* - померять расстояния, определить ближайших
- *логические* - построить правило (комбинацию предикатов)
- *статистические* - восстановить плотность, определить вероятность
- *линейные* - построить разделяющую поверхность
- *композиции* - собрать несколько классификаторов в один

Линейные методы: о задаче классификации

$X = (x, y)$ - датасет

$Y = \{-1, 1\}$ - метки классов

Линейные методы: о задаче классификации

$X = (x, y)$ - датасет

$Y = \{-1, 1\}$ - метки классов

алгоритм классификации

$$a(x, w) = \text{sign}(f(x, w))$$

$f(x, w)$ - дискриминантная ф-ция

w - вектор параметров

Линейные методы: о задаче классификации

$X = (x, y)$ - датасет

$Y = \{-1, 1\}$ - метки классов

разделяющая поверхность

$$f(x, w) = 0$$

алгоритм классификации

$$a(x, w) = \text{sign}(f(x, w))$$

$f(x, w)$ - дискриминантная ф-ция

w - вектор параметров

Линейные методы: о задаче классификации

$X = (x, y)$ - датасет

$Y = \{-1, 1\}$ - метки классов

разделяющая поверхность

$$f(x, w) = 0$$

алгоритм классификации

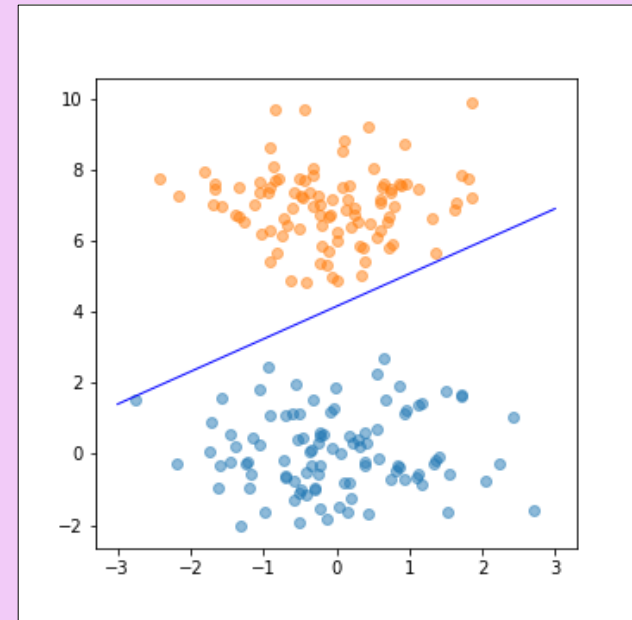
$$a(x, w) = \text{sign}(f(x, w))$$

$f(x, w)$ - дискриминантная ф-ция

w - вектор параметров

пример: линейно разделимые данные
разделяющая поверхность - прямая

$$w_1 \cdot x + w_0 = 0$$



Линейные методы: о задаче классификации

$X = (x, y)$ - датасет

$Y = \{-1, 1\}$ - метки классов

разделяющая поверхность

$$f(x, w) = 0$$

алгоритм классификации

$$a(x, w) = \text{sign}(f(x, w))$$

$f(x, w)$ - дискриминантная ф-ция

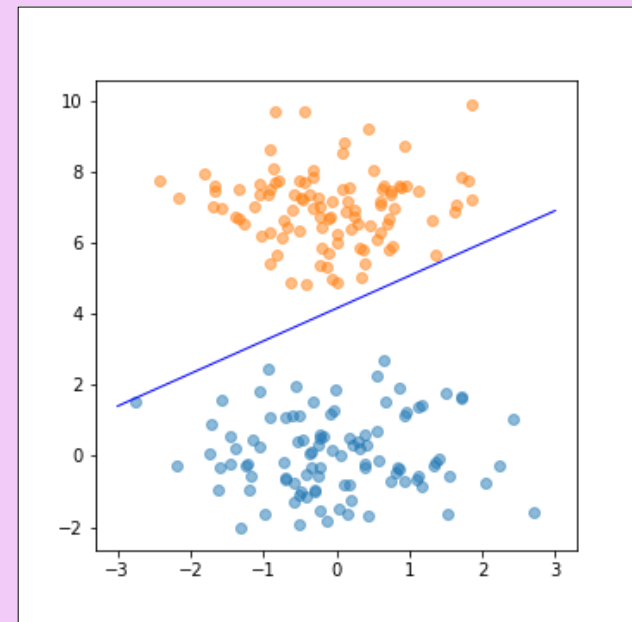
w - вектор параметров

задача:

заданы данные X, Y и вид функции f
найти вектор параметров w ?

пример: линейно разделимые данные
разделяющая поверхность - прямая

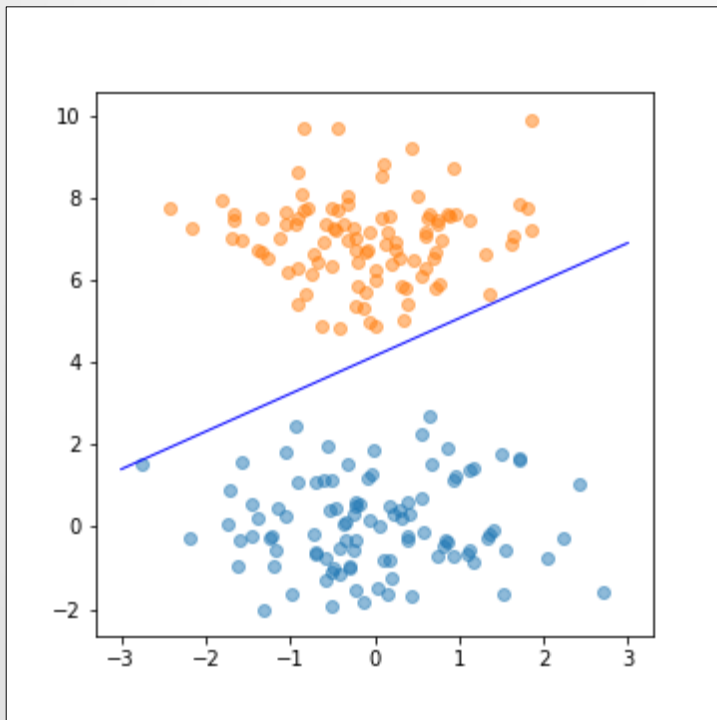
$$w_1 \cdot x + w_0 = 0$$



Линейные методы: отступы

отступ - насколько далеко объект x от разделяющей поверхности

$$M(x, w) = y \cdot f(x, w)$$



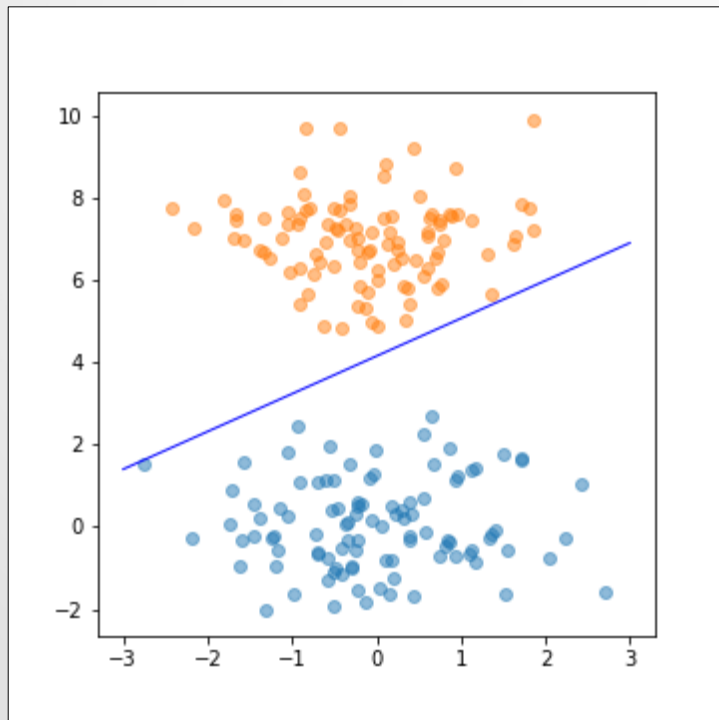
$y \in \{-1, 1\}$ - метка класса

$f(x, w)$ - дискриминантная функция

Линейные методы: отступы

отступ - насколько далеко объект x от разделяющей поверхности

$$M(x, w) = y \cdot f(x, w)$$



$y \in \{-1, 1\}$ - метка класса

$f(x, w)$ - дискриминантная функция

$M(x, w) < 0$ - алгоритм ошибается на x

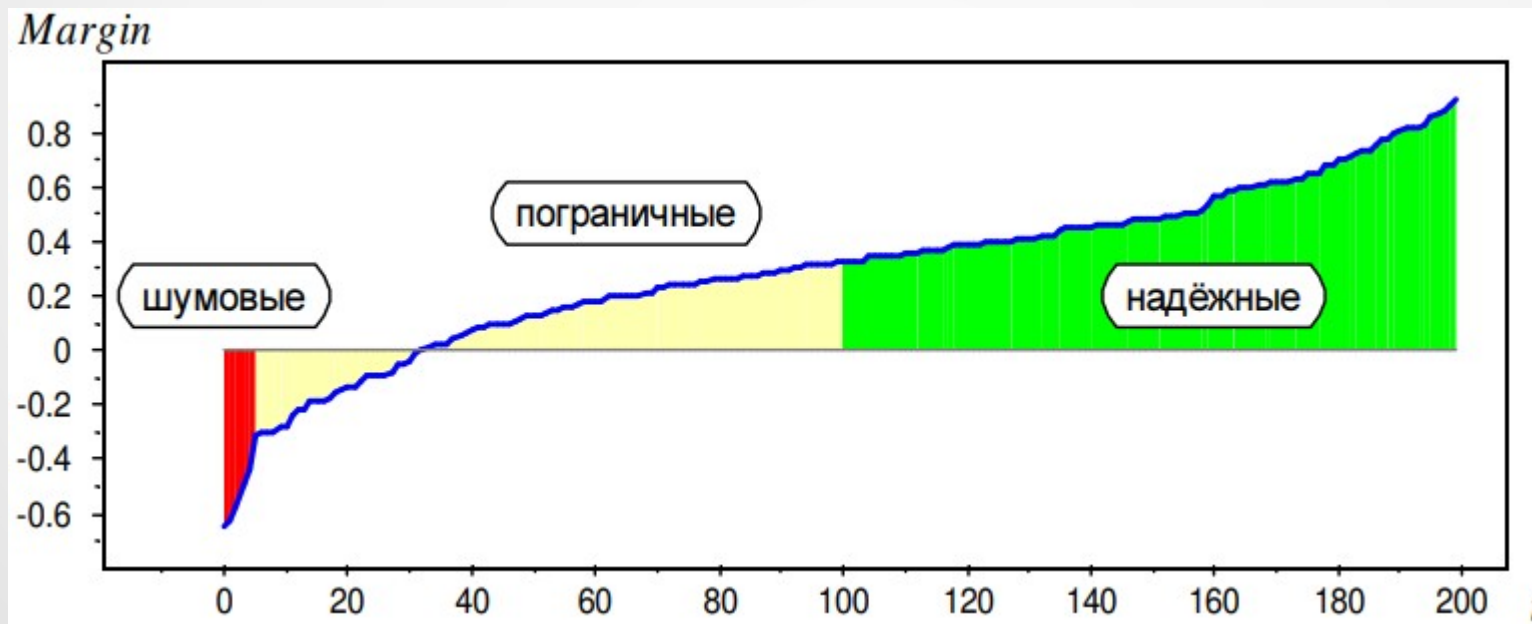
Линейные методы: отступы

отступ - насколько далеко объект от разделяющей поверхности

$$M(x, w) = y \cdot f(x, w)$$

$y \in \{-1, 1\}$ - метка класса

$f(x, w)$ - дискриминантная функция



$M(x, w) < 0$ - алгоритм ошибается на x

Линейные методы: эмпирический риск

функционал эмпирического риска, (число ошибок)

$$Q(x, w) = \sum_x [M(x, w) < 0]$$

$M(x, w) = f(x, w) \cdot y$ - отступ объекта x

$y \in \{-1, 1\}$ - метка класса

$f(x, w)$ - дискриминантная функция

$M(x, w) < 0$ - алгоритм ошибается на x

Линейные методы: функция потерь

функционал эмпирического риска

$$Q(x, w) = \sum_x [M(x, w) < 0]$$

Линейные методы: функция потерь

функционал эмпирического риска

$$Q(x, w) = \sum_x [M(x, w) < 0]$$

[$M < 0$] это пороговая функция,
не учитываем значение отступа M ,
оптимизировать не удобно,
заменим её...

Линейные методы: функция потерь

функционал эмпирического риска

$$Q(x, w) = \sum_x [M(x, w) < 0]$$

[$M < 0$] это пороговая функция,

не учитываем значение отступа M ,

оптимизировать не удобно,

заменим её...

построим аппроксимацию Q

определим **функцию потерь $L(M)$** - невозрастающая, неотрицательная

$$\tilde{Q}(x, w) = \sum_x L(M(x, w)) \rightarrow \min$$

$$Q(x, w) \leq \tilde{Q}(x, w)$$

Линейные методы

функционал эмпирического риска

$$Q(x, w) = \sum_x [M(x, w) < 0]$$

$[M < 0]$ это пороговая функция, оптимизировать не удобно, заменим её...

варианты для замены $[M < 0]$

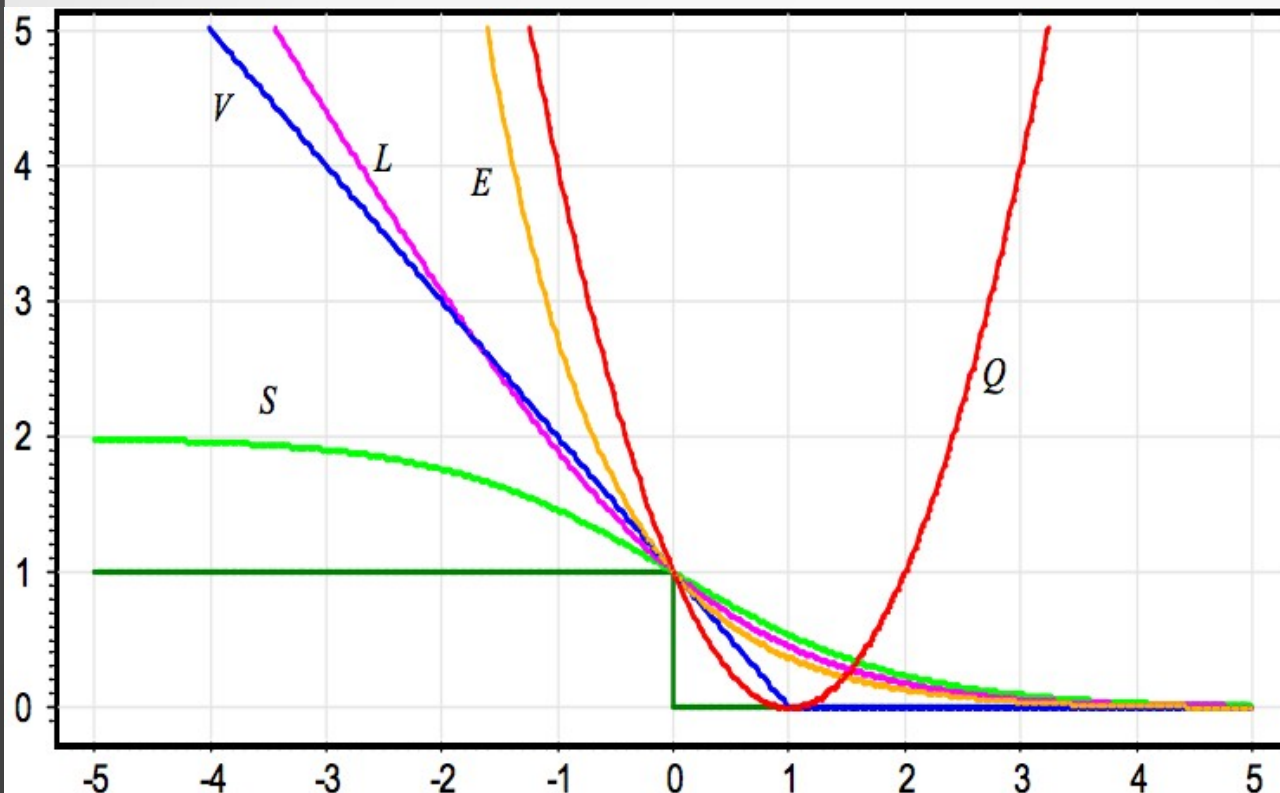
$$L(M) = \log_2 \left(1 + \frac{1}{\exp(M)} \right) \quad \text{логарифмическая}$$

$$V(M) = (1 - M)_+ \quad \text{кусочно-линейная}$$

$$Q(M) = (1 - M)^2 \quad \text{квадратичная}$$

$$E(M) = \frac{1}{\exp(M)} \quad \text{экспоненциальная}$$

$$S(M) = \frac{1}{2 \cdot (1 + \exp(M))} \quad \text{сигмоид}$$



Линейные методы: линейный классификатор

рассмотрим линейный классификатор,

дискриминантная функция $f(x, w)$ это гиперплоскость

$$f(x, w) = \sum_{i=1}^n x_i \cdot w_i - w_0 \quad - \text{дискриминантная функция}$$

$$a(x, w) = \text{sign}(f(x, w)) = \text{sign}\left(\sum_{i=1}^n x_i \cdot w_i - w_0\right) = \text{sign}(\langle x, w \rangle)$$

$$M(x, w) = \langle x, w \rangle \cdot y \quad - \text{отступ на объекте } \mathbf{x} \text{ класса } \mathbf{y}$$

$$\text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

Линейные методы

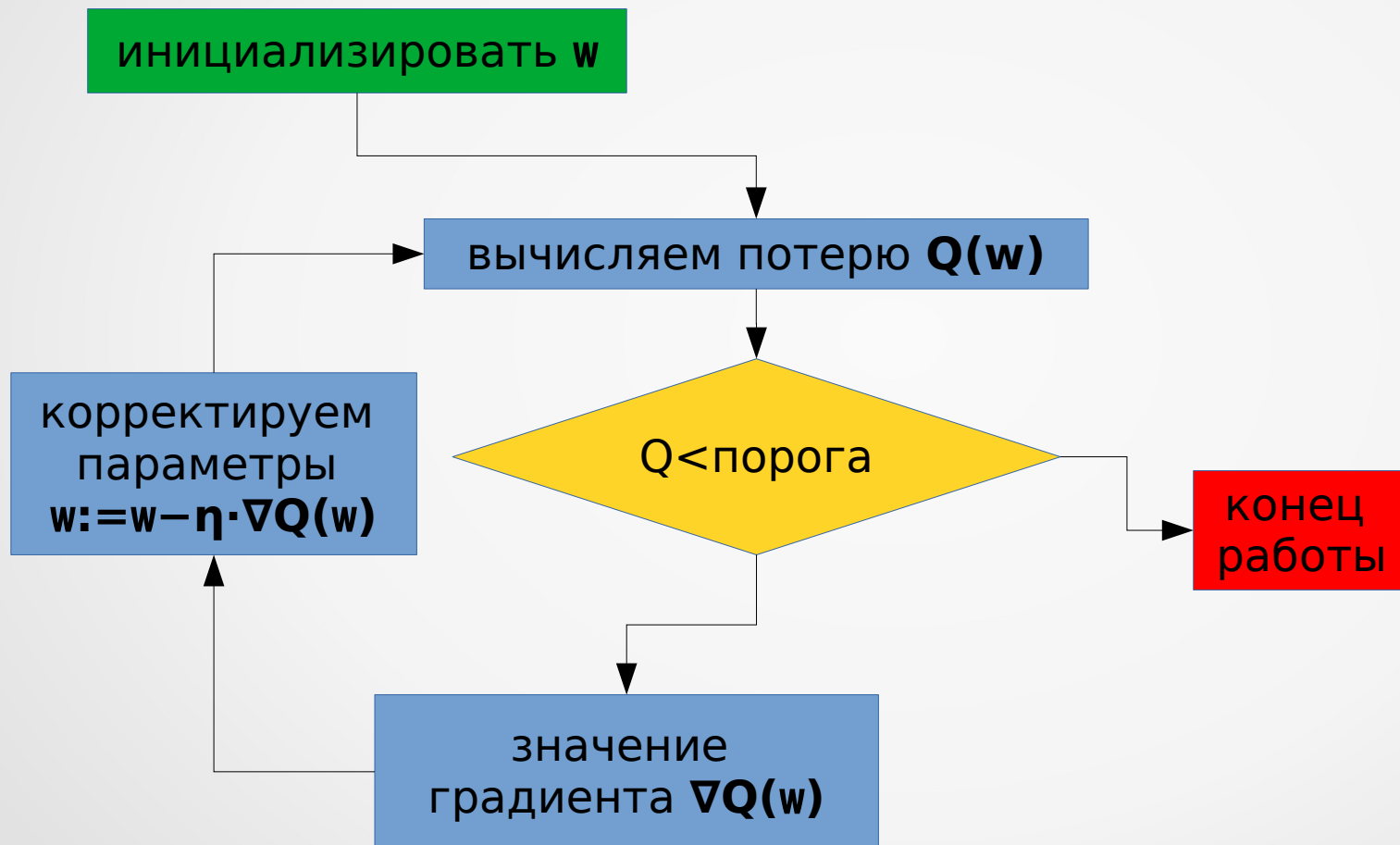
обучение классификатора как задача оптимизации

$$Q(w; X) = \sum_{x \in X} L(\langle x, w \rangle \cdot y) \rightarrow \min_w$$

можно использовать градиентные методы

$$\nabla Q(w) = \left(\frac{\partial Q(w)}{\partial w_j} \right)_{j=0}^n \text{ - вектор градиента ф-ции } Q$$

Линейные методы: градиентный спуск (GD)



Линейные методы: стохастический градиентный спуск (SGD)

инициализировать w

вычисляем суммарную
потерю $Q(w)$ на X

$Q < \text{порога}$

конец
работы

корректируем
суммарную потерю
 $Q := \lambda Q_j + (1 - \lambda)Q$

выбираем
случайный x_j

вычисляем значение
градиента $\nabla Q(w, x_j)$

вычисляем
потерю для объекта x_j
 $Q_j = Q(w, x_j)$

корректируем
параметры
 $w := w - \eta \cdot \nabla Q(w, x_j)$

Линейные методы

«зоопарк» методов

- вид разделяющей поверхности $f(x, w)$
(линейная, нелинейная)
- вид функции потерь $L(M)$
- вид метода оптимизации $Q(w) \rightarrow \min$

Линейные методы: итог

- линейные методы строят разделяющие поверхности в пространстве признаков
- использования нелинейных поверхностей позволяет разделять линейно неразделимые наборы
- аппроксимация пороговой ф-ции потерь позволяет использовать градиентные методы оптимизации
- метод стохастического градиента SGD подходит для обучения на больших данных

Линейные методы: литература

git clone https://github.com/mechanoid5/ml_lectorium.git

К.В. Воронцов Линейный классификатор и стохастический градиент.
<http://www.machinelearning.ru>