KARPOV.COURSES >>>
КОНСПЕКТ

Конспект > 13 урок > Метод опорных векторов

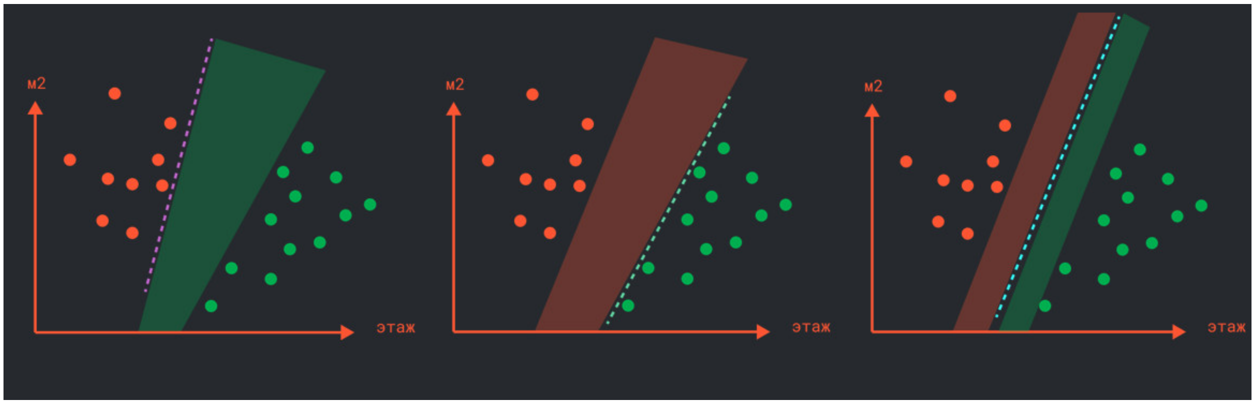
>Оглавление

>Метод опорных векторов(SVM)

Метод опорных векторов(SVM - Support Vector Machine) - способ построения модели классификации, суть которого заключается в построении разделяющей гиперплоскости.

Однако в отличии логистической регрессии, рассмотренной ранее, данный алгоритм работает в предположении, что чем больше расстояние (зазор) между разделяющей гиперплоскостью и объектами разделяемых классов, тем меньше будет средняя ошибка классификатора.

При этом гиперплоскость должна пройти так, чтобы расстояние до ближайшего объекта было максимальным.



То есть наша цель максимизация расстояния до ближайшего объекта. Математически это можно записать следующим образом:

$$\min_{x \in X} \quad \rho(x_i, \beta) -> max_{\beta}$$

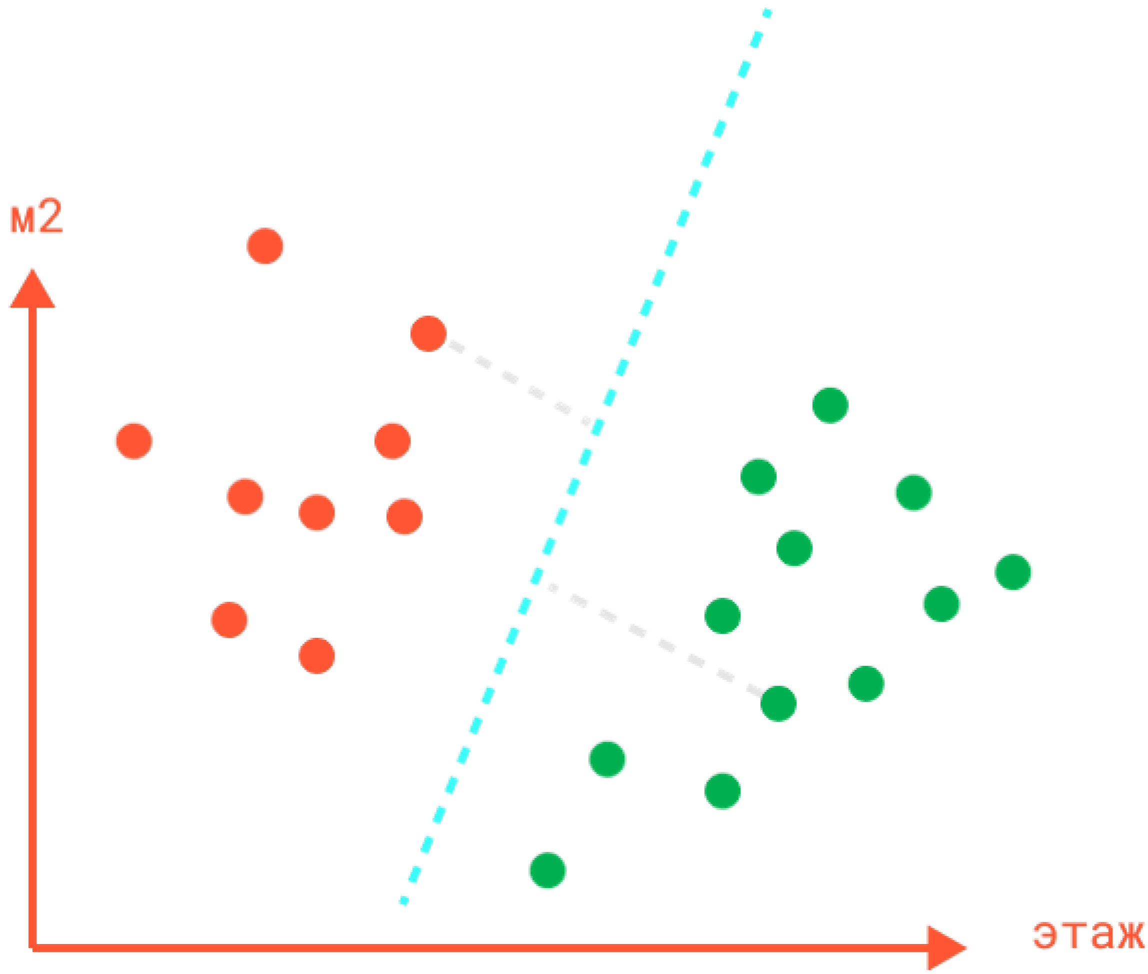
$s.t.$ (при условии)

$$y_i \cdot \langle x_i, \beta \rangle \geq 0$$

>Расстояние от точки до плоскости

Чтобы найти расстояние от точки до плоскости, нужно разделить модуль скалярного произведения на модуль(длину) вектора коэффициентов β :

$$\rho\langle x_i, \beta\rangle = \frac{|\langle \beta, x_i \rangle|}{|\beta|} = \frac{|\beta_1 \cdot d_1 + \beta_2 \cdot d_2 + \ldots + \beta_n \cdot d_n + \beta_0|}{\sqrt{(\beta_1^2 + \beta_2^2 + \ldots + \beta_n^2)}}$$



Предположим, мы построили следующее уравнение гиперплоскости для нашей выборки:

$a(x) = sgn(1 \cdot \text{м}^2 + 1.5 \cdot \text{этаж} - 30)$

Допустим имеем два объекта:

$x_1 = (10, 6)$

$x_2 = (19, 11)$

Подставим эти значения в уравнение, чтобы найти расстояние до разделяющей гиперплоскости:

$\rho\langle x_1, \beta^* \rangle = \frac{|1 \cdot 10 + 1.5 \cdot 6 - 30|}{\sqrt{(1^2 + 1.5^2)}} \approx 6.1$

$$\rho\langle x_2, \beta^*\rangle = \frac{|1 \cdot 19 + 1.5 \cdot 11 - 30|}{\sqrt{(1^2 + 1.5^2)}} \approx 3.05$$

Можем сделать вывод, что расстояние для первого объекта до гиперплоскости в два раза больше, чем для второго.

>Уравнение бинарной классификации

Факт: Гиперплоскость нейтральна к умножению

Если все коэффициенты разделяющей гиперплоскости умножить на какое-либо одинаковое неотрицательное нулевое значение, то гиперлоскость никак не изменится.

Пользуясь этим свойством, мы можем упростить формулу расчета минмального расстояния до гиперплоскости:

$$\min_{x \in X} \langle \beta, X \rangle = 1$$

То есть будем так перевзвешивать коэффициенты β, чтобы скалярное произведение векторов было равным единице.

Тогда уравнение бинарной классификации можно преобразовать следующим образом:

$$\min_{x \in X} \frac{|\langle \beta, X \rangle|}{|\beta|} -> max_{\beta}$$

$$s.t.$$

$$y_i \cdot \langle x_i, \beta \rangle \geq 0$$

$$\min_{x \in X} \langle \beta, X \rangle = 1$$

$$\min_{x \in X} \frac{|\langle \beta, X \rangle|}{|\beta|} = \frac{\min_{x \in X}}{|\langle \beta, X \rangle|} = \frac{1}{|\beta|}$$

Итак наше выражение бинарной классификации принимает вид:

$$\frac{1}{|\beta|} —> max_{\beta}$$

$$s.t.$$

$$y_i \cdot \langle x_i, \beta \rangle \geq 0$$

$$\min_{x \in X} \langle \beta, X \rangle = 1$$

Далее, логично предположить, что чем меньше длина вектора коэффициентов β, тем больше $\frac{1}{|\beta|}$, а значит задачу снова можно свести к минимизации!

Для этого снова будем считать производную, но для начала еще упростим выражение длины вектора коэффициентов, избавившись от корня:

$$|\beta| = \sqrt{(\beta_1^2 + \beta_2^2 + ... + \beta_n^2)}$$

Для этого просто возведем данное выражение квадрат, так как функция минимизации монотонна, то это никак не повлияет на ее результат.

Также упростим условия для скалярного произведения векторов. Зная, что

$$\min_{x \in X} \langle \beta, X \rangle = 1$$

$$y_i \cdot \langle x_i, \beta \rangle \geq 0$$

$y_i \in -1, 1$, объединим эти выражения в одно:

$$y_i \cdot \langle x_i, \beta \rangle \geq 1$$

В итоге мы получили финальную формул бинарной классификации:

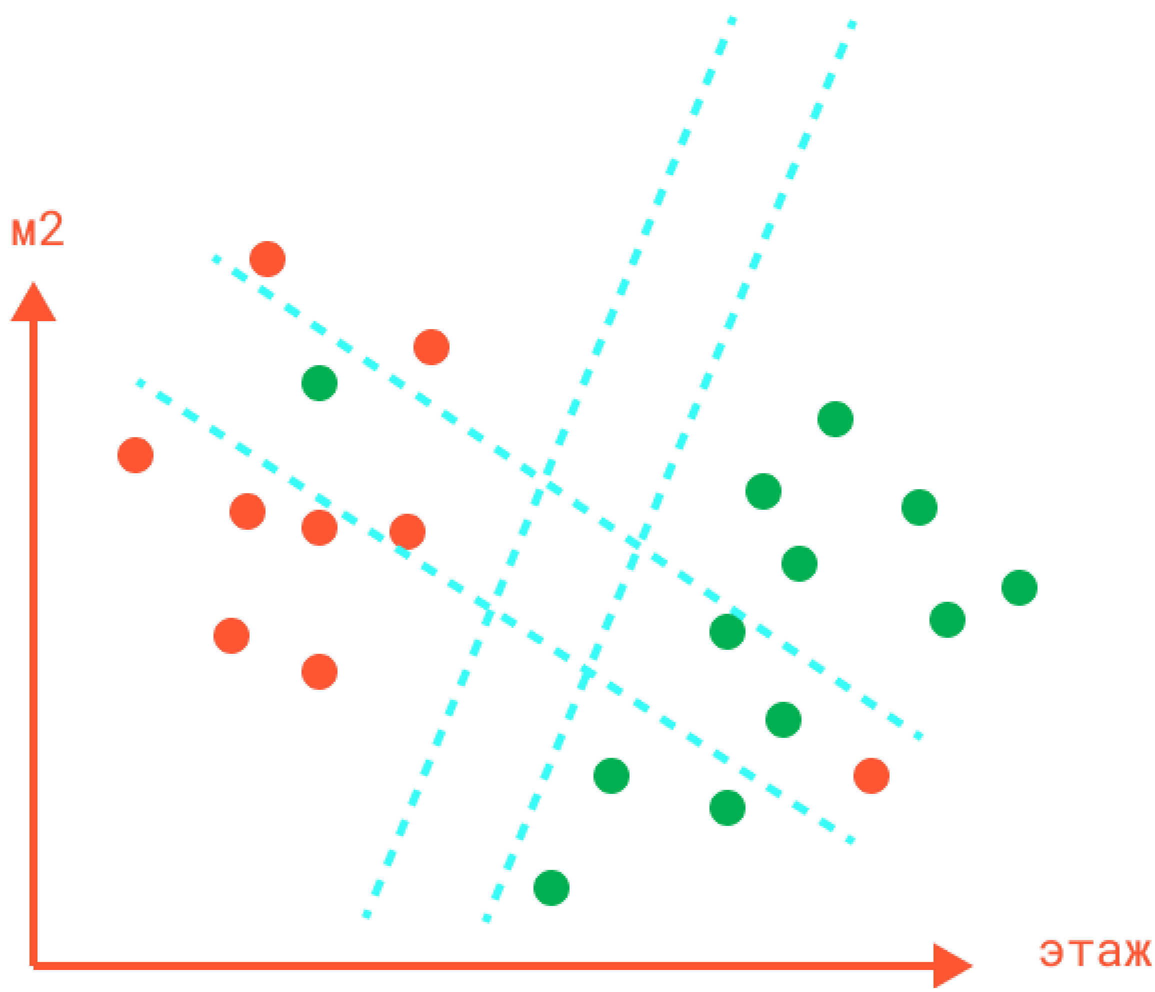
$$|\beta|^2 -> min_{\beta}$$

$$s.t.$$

$$y_i \cdot \langle x_i, \beta \rangle \geq 1$$

>Линейная неразделимость и регуляризация

Не всегда значения признаков распределены так, что между ними можно провести разделяющую гиперплоскость.



Проблема построения гиперплоскости для линейно-неразделимых объектов в том, что условием ее построения является наличие положительного отступа, то есть невозможность допущения ошибки ни на одном объекте.

Давайте смягчим это условие и введем дополнительный параметр ξ, теперь требование к отступам таково, что чтобы отступы для объеков были не меньше, чем некое число $1 - \xi$. Этот параметр можно подбирать.

Однако делать это для каждого объекта кажется довольно трудозатратным. Давайте попробуем автоматизировать подбор подходящих \xi используя уже известные нам методы регуляризации.

Для этого добавим к функции минимизации \beta сумму параметров ξ, умноженную на коэффициент регуляризации:

$$|\beta|^2 + \lambda \cdot \sum_i^n \xi_i -> min_{\beta,\xi}$$

$$s.t.$$

$$y_i \cdot \langle x_i, \beta \rangle \geq 1 - \xi_i$$

$$\xi_i \geq 0$$

Такая модель будет штрафовать за слишком большие параметры ξ, ее строгость можно регулировать настраивая коэффициент λ

Сравнение регуляризации разной силы

И снова попробуем упростить полученное условное выражение

$y_i \cdot \langle x_i, \beta \rangle \geq 1 - \xi_i$

$\xi_i \geq 0$

Перепишем в следующем виде:

$\xi_i \geq 1 - y_i \cdot \langle x_i, \beta \rangle$

$\xi_i \geq 0$

Преобразуем в равенство и получаем:

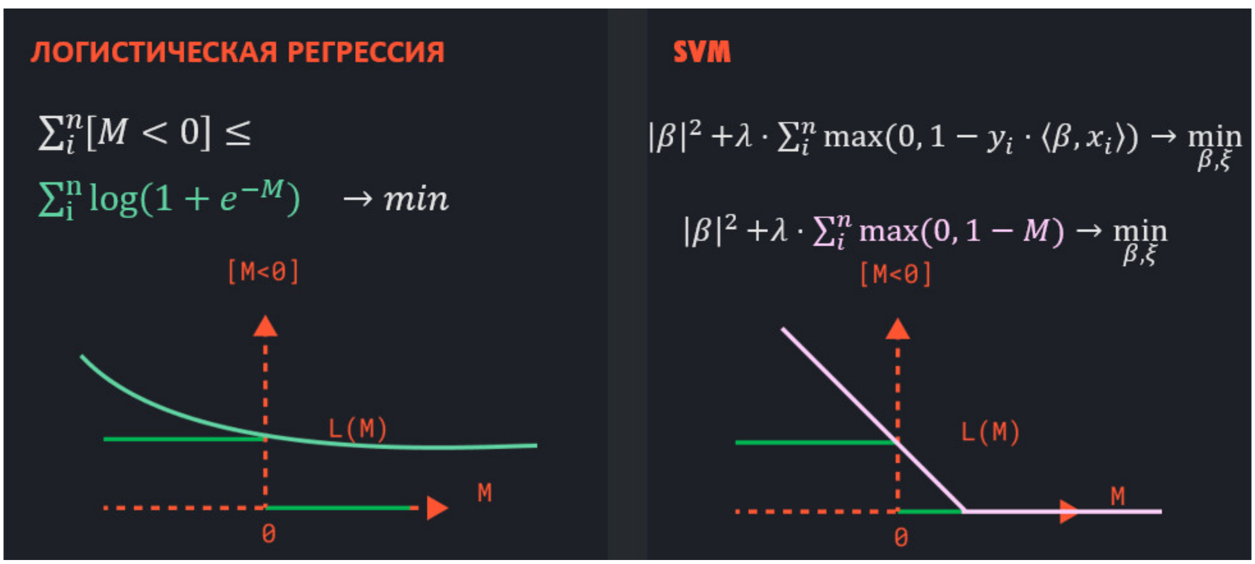
$\xi_i = max(0, 1 - y_i \cdot \langle x_i, \beta \rangle)$

Подставим выражение в целевую функцию:

$|\beta|^2 + \lambda \cdot \sum_i^n max(0, 1 - y_i \cdot \langle x_i, \beta \rangle) {-}{-}> min_{\beta,\xi}$

>SVM vs Логистическая регрессия

Сравним данный подход с методом логистической регресии



- SVM является частным случаем минимизации индикаторов отступов со специфической верхней оценкой и модификацией в виде регуляризации.
- В отличии от логистической регрессии, он не учитывает вероятности и не стремится к максимизации уверенности в прогнозе.
- Его основная задача - построение разделяющей гиперплоскости с достаточно широкой полосой.
- Функция потерь для SVM носит название hinge loss
- При этом мы все еще можем оценить вероятность с помощью метода калибровки.