Лекция 6
Линейные модели классификации. Часть 2.

Кантонистова Е.О.

ВШЭ, 2021

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Хотим предсказывать не классы, а вероятности классов.

- Линейная регрессия: $a(x, w) = (x, w) \in \mathbb{R}$
- Логистическая регрессия: $\boldsymbol{a(x, w) = g((x, w))}$,

где $g(z) = \frac{1}{1+e^{-z}}$ - сигмоида (логистическая функция)

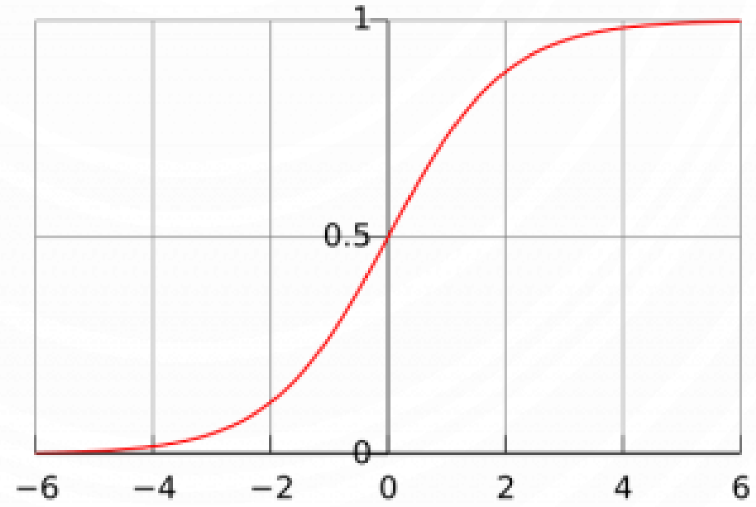
ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Хотим предсказывать не классы, а вероятности классов.

- Линейная регрессия: $a(x, w) = (x, w) \in \mathbb{R}$
- Логистическая регрессия: $\boldsymbol{a(x, w) = g((x, w))}$,

где $g(z) = \frac{1}{1+e^{-z}}$ - сигмоида (логистическая функция),

$g(z) \in (0; 1)$.



Логистическая регрессия: $\boldsymbol{a(x, w) = \frac{1}{1+e^{-(x,w)}}}$

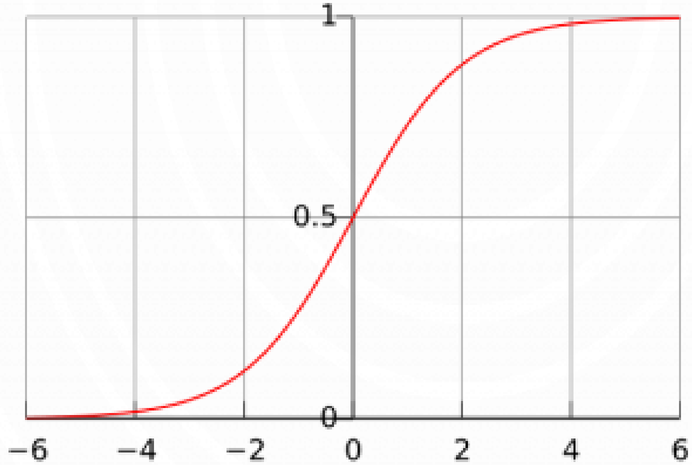
ВЕРОЯТНОСТНЫЙ СМЫСЛ

- $a(x, w)$ – вероятность того, что $y = +1$ на объекте x, то есть

$$\boldsymbol{a(x, w) = P(y = +1 | x; w)}$$

РАЗДЕЛЯЮЩАЯ ГРАНИЦА

Предсказываем $y = +1$, если $a(x, w) \geq 0.5$.



$a(x, w) = g((x, w)) \geq 0.5$, если $(x, w) \geq 0$.

Получаем, что

- $y = +1 \text{ при } (x, w) \geq 0$
- $y = -1 \text{ при } (x, w) < 0$,

т.е. $(x, w) = 0$ – разделяющая гиперплоскость.

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Логистическая регрессия - это линейный классификатор!

ФУНКЦИЯ ПОТЕРЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

Если взять квадратичную функцию потерь

$L(a, y) = (a - y)^2$,

то возникнут проблемы:

- $Q(a, X) = \frac{1}{l} \sum_{i=1}^{l} \left(\frac{1}{1 + e^{-(x,w)}} - y \right)^2$ - не выпуклая функция (*можем не попасть в глобальный минимум при оптимизации*)

ФУНКЦИЯ ПОТЕРЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

Если взять квадратичную функцию потерь

$L(a, y) = (a - y)^2$,

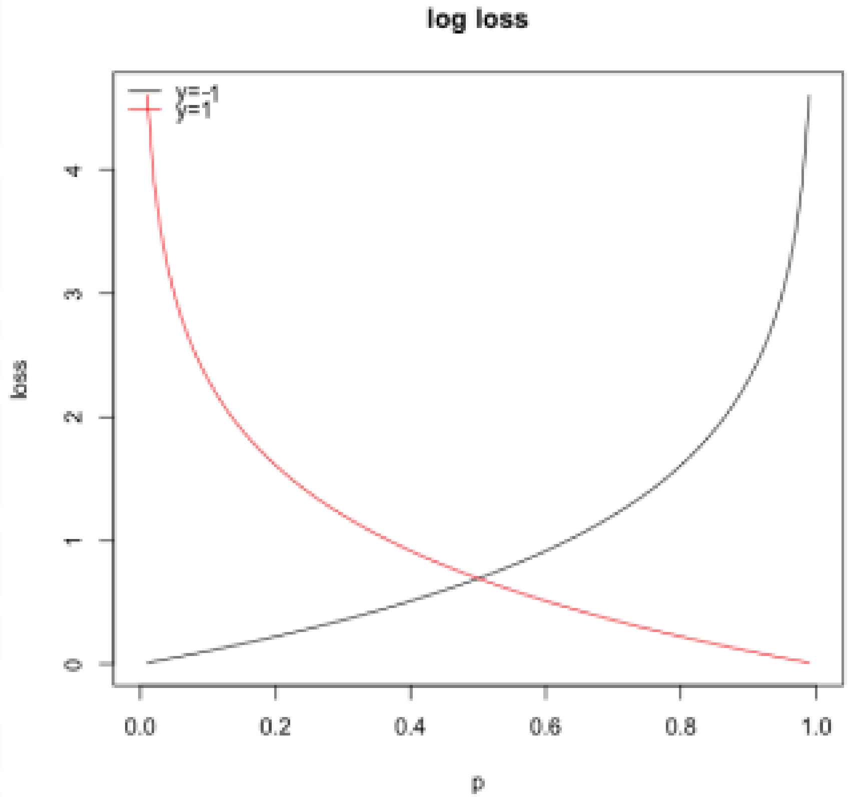
то возникнут проблемы:

- $Q(a, X) = \frac{1}{l}\sum_{i=1}^{l}\left(\frac{1}{1+e^{-\langle x,w\rangle}} - y\right)^2$ - не выпуклая функция (*можем не попасть в глобальный минимум при оптимизации*)
- *На совсем неправильном предсказании маленький штраф* (пусть предсказали вероятность 0% на объекте класса $y = +1$, тогда штраф всего $(1 - 0)^2 = 1$)

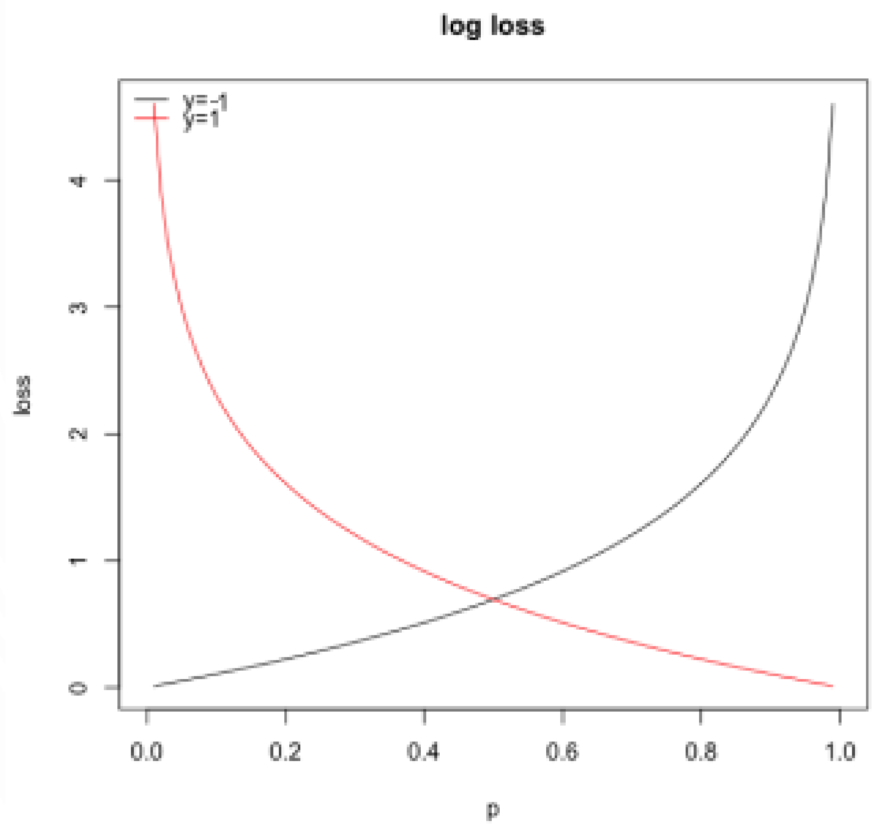
ФУНКЦИЯ ПОТЕРЬ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

Возьмем логистическую функцию потерь (log-loss):

$$Q(w) = -\sum_{i=1}^{l}\left([y_i = +1] \cdot \log\left(a(x_i, w)\right) + [y_i = -1] \cdot \log\left(1 - a(x_i, w)\right)\right)$$



ЛОГИСТИЧЕСКАЯ ФУНКЦИЯ ПОТЕРЬ

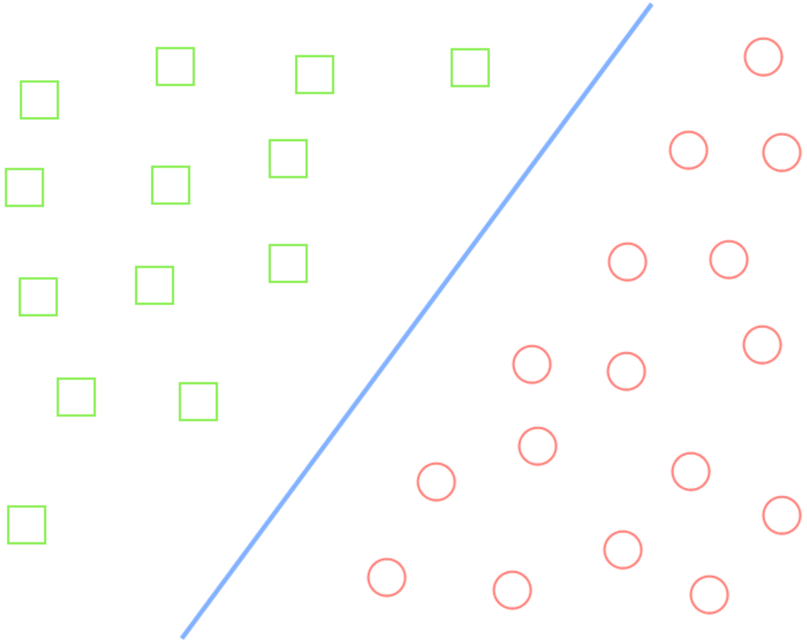


- если $a(x, w) = 1$ и $y = +1$, то штраф $L(a, y) = 0$
- если $a(x, w) \to 0$, а $y = +1$, то штраф $L(a, y) \to +\infty$

МЕТОД ОПОРНЫХ ВЕКТОРОВ

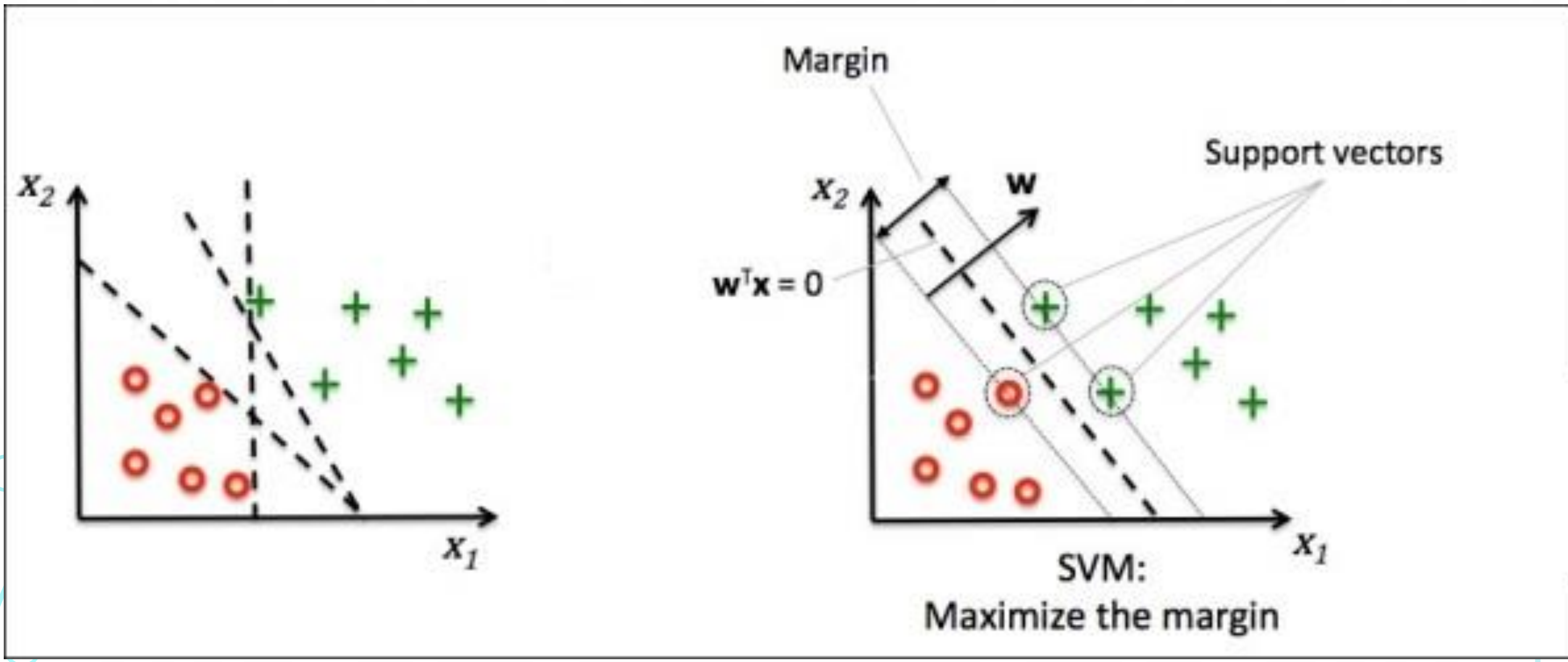
ЛИНЕЙНО РАЗДЕЛИМАЯ ВЫБОРКА

Выборка ***линейно разделима***, если существует такой вектор параметров w^*, что соответствующий классификатор $a(x)$ не допускает ошибок на этой выборке.



Метод опорных векторов: разделимый случай

Цель метода опорных векторов (Support Vector Machine) – максимизировать ширину разделяющей полосы.



МЕТОД ОПОРНЫХ ВЕКТОРОВ: РАЗДЕЛИМЫЙ СЛУЧАЙ

- $a(x) = sign\big((w, x) + w_0\big)$
- Нормируем параметры w и w_0 так, что

$$\min_{x \in X} |(w, x) + w_0| = 1$$

Метод опорных векторов: разделимый случай

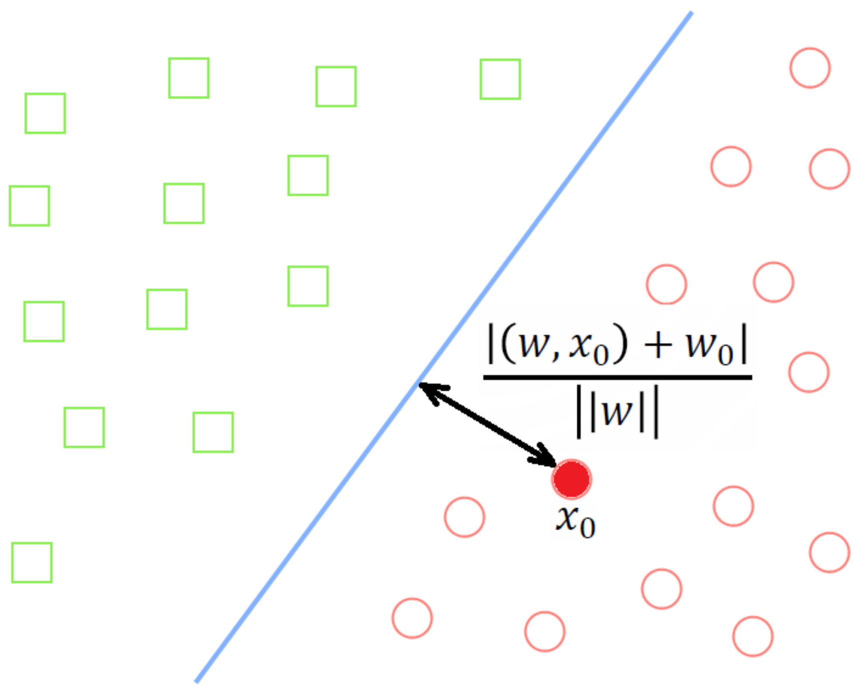
- $a(x) = sign\big((w, x) + w_0\big)$
- Нормируем параметры w и w_0 так, что

$$\min_{x \in X} |(w, x) + w_0| = 1$$

Расстояние от точки x_0 до разделяющей гиперплоскости, задаваемой

классификатором:

$$\rho(x_0, a) = \frac{|(w, x_0) + w_0|}{||w||}$$



МЕТОД ОПОРНЫХ ВЕКТОРОВ: РАЗДЕЛИМЫЙ СЛУЧАЙ

- Нормируем параметры w и w_0 так, что

$$\min_{x \in X} |(w, x) + w_0| = 1$$

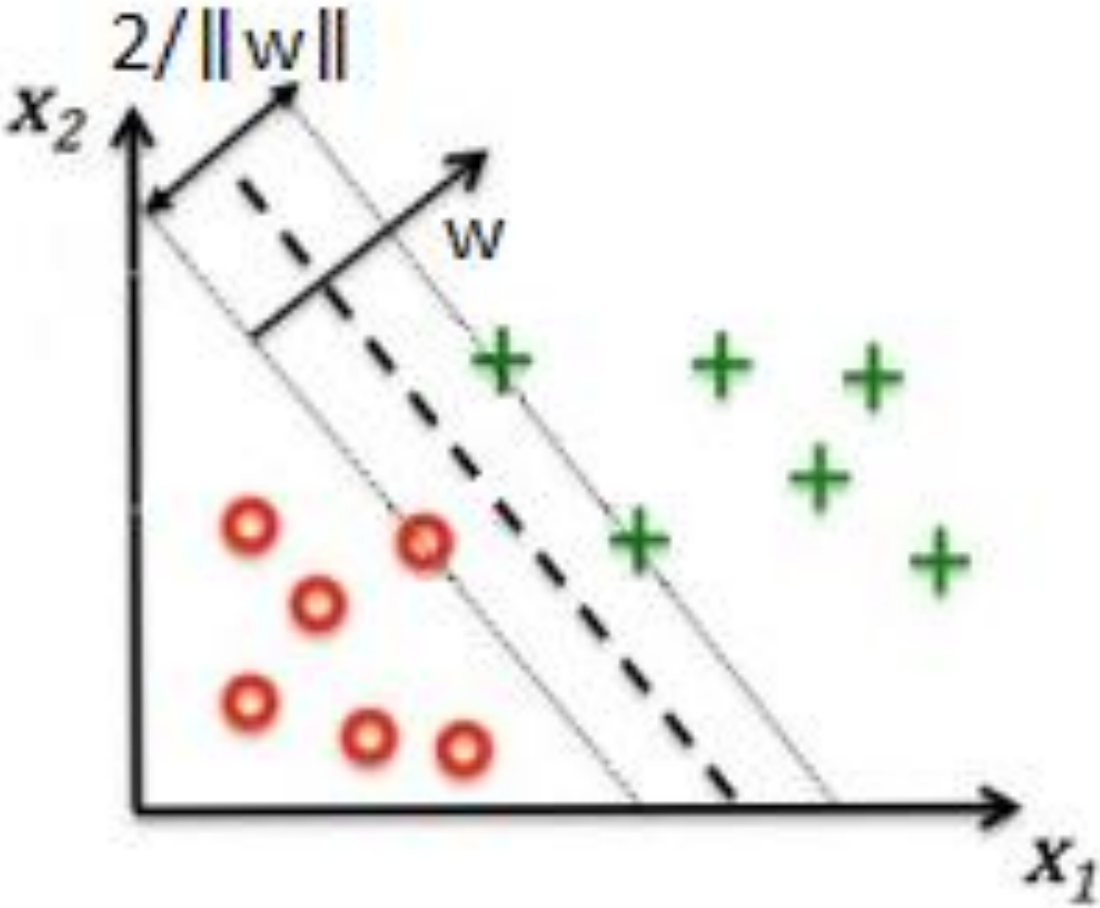
Тогда расстояние от точки x_0 до разделяющей гиперплоскости, задаваемой классификатором:

$$\rho(x_0, a) = \frac{|(w, x_0) + w_0|}{||w||}$$

- Расстояние до ближайшего объекта $x \in X$:

$$\min_{x \in \mathrm{X}} \frac{|(w, x) + w_0|}{||w||} = \frac{1}{||w||} \min_{x \in X} |(w, x) + w_0| = \frac{\mathbf{1}}{||\boldsymbol{w}||}$$

РАЗДЕЛЯЮЩАЯ ПОЛОСА



Оптимизационная задача SVM для разделимой выборки

$$\begin{cases} \frac{1}{2}||w||^2 \to \min\limits_{w} \\ y_i\big((w, x_i) + w_0\big) \geq 1, i = 1, \dots, l \end{cases}$$

Утверждение. Данная оптимизационная задача имеет единственное решение.

ЛИНЕЙНО НЕРАЗДЕЛИМАЯ ВЫБОРКА

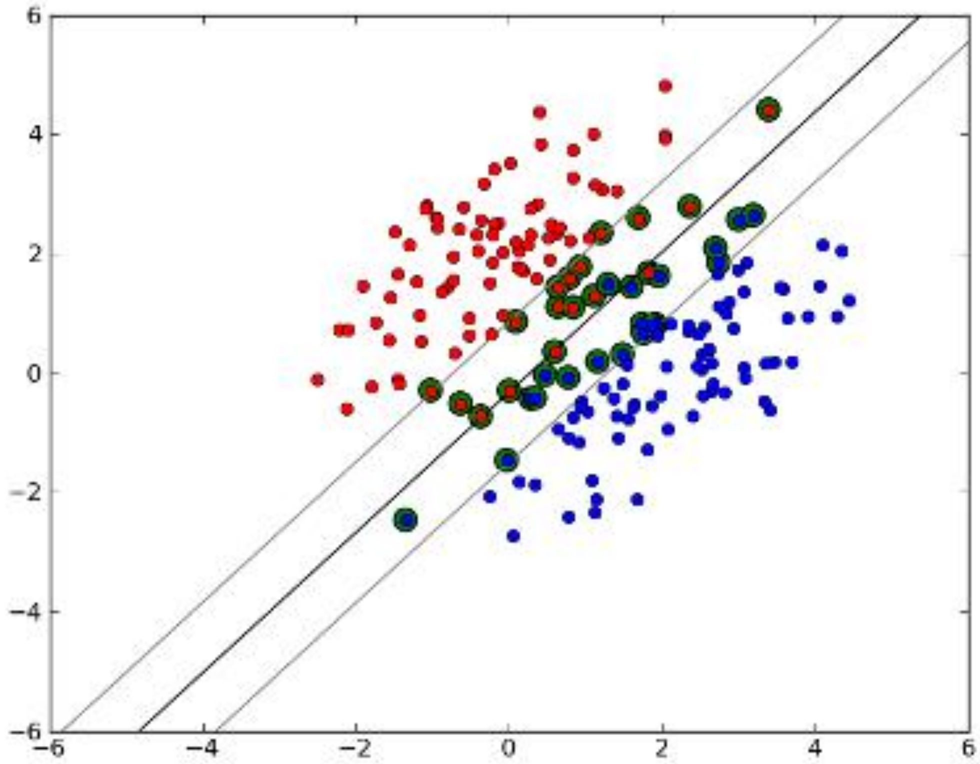
- Существует хотя бы один объект $x \in X$, что

$$y_i\big((w, x_i) + w_0\big) < 1$$

ЛИНЕЙНО НЕРАЗДЕЛИМАЯ ВЫБОРКА

- Существует хотя бы один объект $x \in X$, что

$$y_i\big((w, x_i) + w_0\big) < 1$$



ЛИНЕЙНО НЕРАЗДЕЛИМАЯ ВЫБОРКА

- Существует хотя бы один объект $x \in X$, что

$$y_i\big((w, x_i) + w_0\big) < 1$$

Смягчим ограничения, введя штрафы $\xi_i \geq 0$:

$$y_i\big((w, x_i) + w_0\big) \geq 1 - \xi_i, i = 1, \dots, l$$

Метод опорных векторов: неразделимый случай

Хотим:

- Минимизировать штрафы $\sum_{i=1}^{l} \xi_i$
- Максимизировать отступ $\frac{1}{||w||}$

Метод опорных векторов: неразделимый случай

Хотим:

- Минимизировать штрафы $\sum_{i=1}^{l} \xi_i$
- Максимизировать отступ $\frac{1}{||w||}$

Задача оптимизации:

$$\begin{cases} \frac{1}{2}\left||w|\right|^2 + C\sum_{i=1}^{l} \xi_i \to \min\limits_{w, w_0, \xi_i} \\ y_i\big((w, x_i) + w_0\big) \geq 1 - \xi_i, i = 1, \dots, l \\ \xi_i \geq 0, i = 1, \dots, l \end{cases}$$

Метод опорных векторов: неразделимый случай

Утверждение. Задача

$$\begin{cases} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \to \min\limits_{w, w_0, \xi_i} \\ y_i\big((w, x_i) + w_0\big) \geq 1 - \xi_i, i = 1, \dots, l \\ \xi_i \geq 0, i = 1, \dots, l \end{cases}$$

Является выпуклой и имеет единственное решение.

СВЕДЕНИЕ К БЕЗУСЛОВНОЙ ЗАДАЧЕ

$$\begin{cases} \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\xi_i \to \min_{w,w_0,\xi_i} (1) \\ y_i\big((w,x_i)+w_0\big) \geq 1-\xi_i, i=1,\dots,l \ (2) \\ \xi_i \geq 0, i=1,\dots,l \ (3) \end{cases}$$

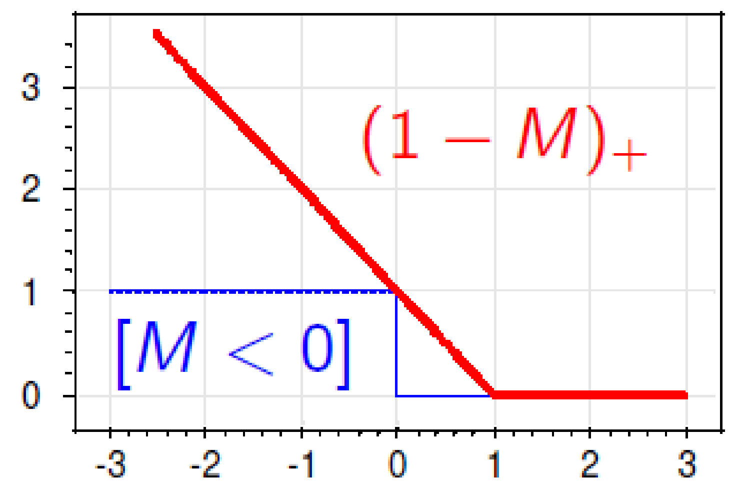
Эту задачу можно переписать в другом, более простом виде:

$$\frac{1}{2}||w||^2 + C\sum_{i=1}^{l}\max\Big(0, 1 - y_i\big((w,x_i)+w_0\big)\Big) \to \min_{w,w_0}$$

МЕТОД ОПОРНЫХ ВЕКТОРОВ: ЗАДАЧА ОПТИМИЗАЦИИ

- На задачу оптимизации SVM можно смотреть, как на *оптимизацию функции потерь* $L(M) = max(0{,}1 - M) = (1 - M)_+$ *с регуляризацией*:

$$Q(a, X) = \sum_{i=1}^{l} \left(1 - M_i(w, w_0)\right)_+ + \frac{1}{2C} ||w||^2 \to \min_{w, w_0}$$

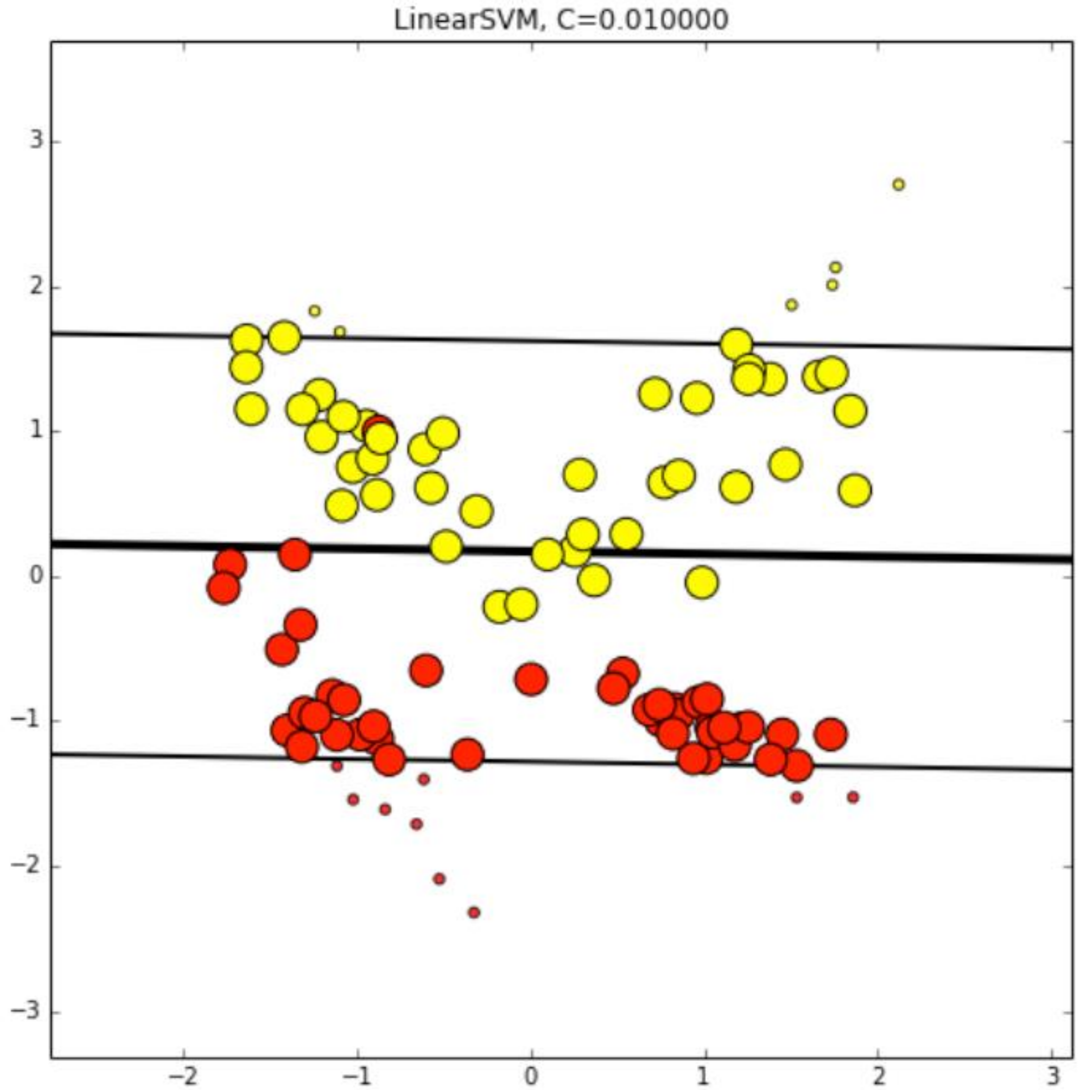


ЗНАЧЕНИЕ КОНСТАНТЫ C

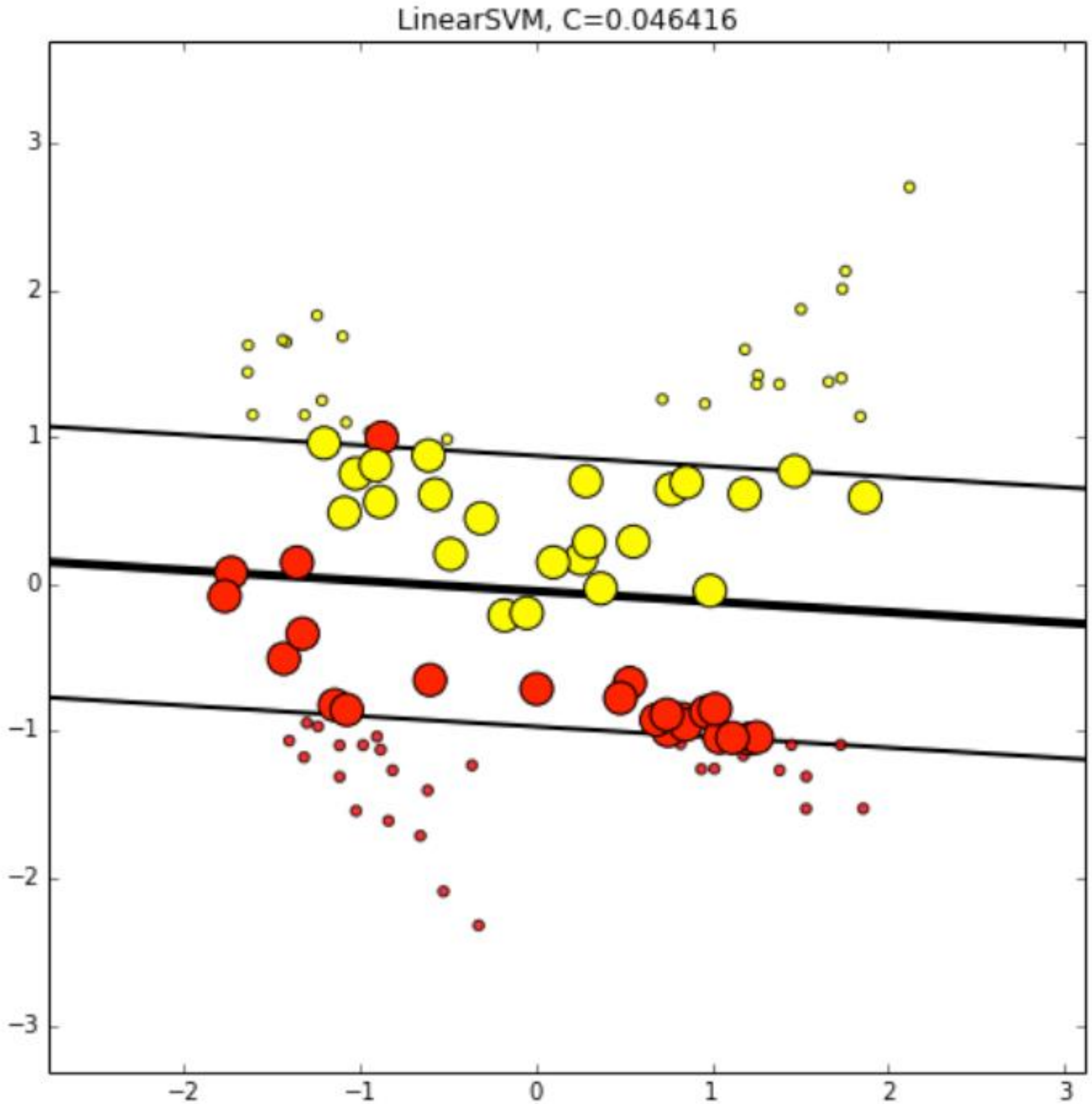
$$\begin{cases} \frac{1}{2}||w||^2 + C\sum_{i=1}^{l} \xi_i \to \min_{w, w_0, \xi_i} (1) \\ y_i\big((w, x_i) + w_0\big) \geq 1 - \xi_i, i = 1, \dots, l\ (2) \\ \xi_i \geq 0, i = 1, \dots, l\ (3) \end{cases}$$

Положительная константа C является управляющим параметром метода и позволяет находить компромисс между максимизацией разделяющей полосы и минимизацией суммарной ошибки.

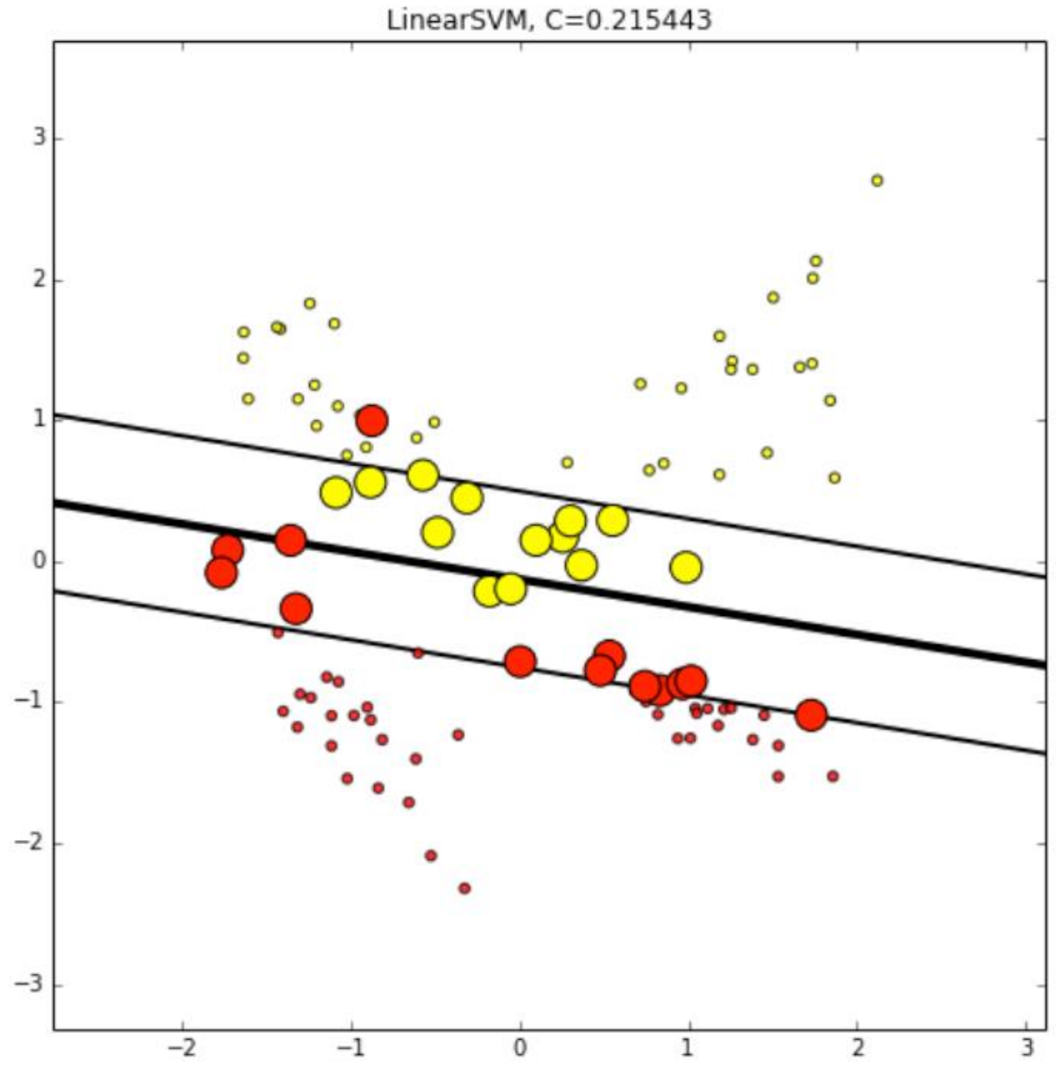
ЗНАЧЕНИЕ КОНСТАНТЫ С



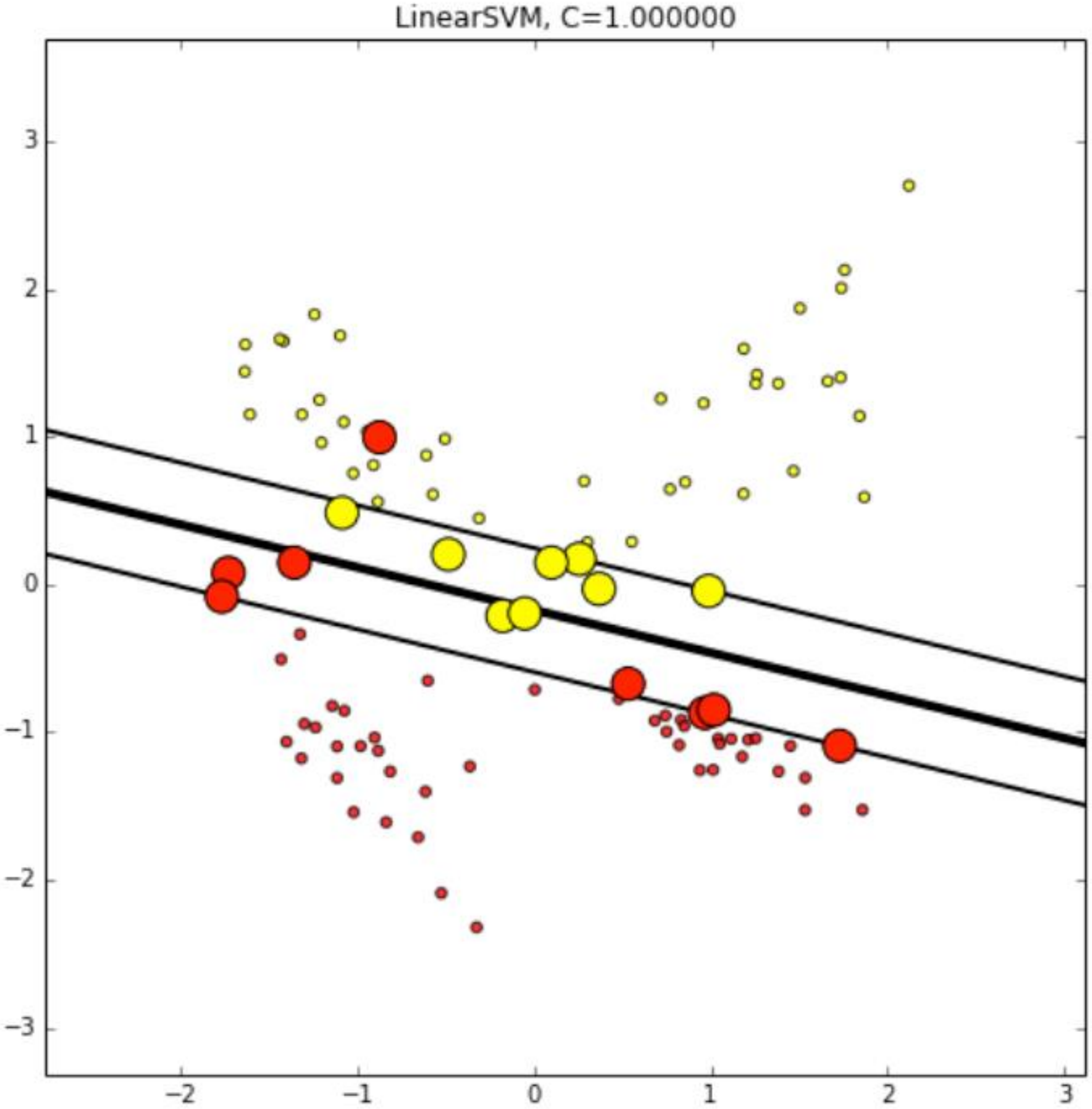
ЗНАЧЕНИЕ КОНСТАНТЫ С



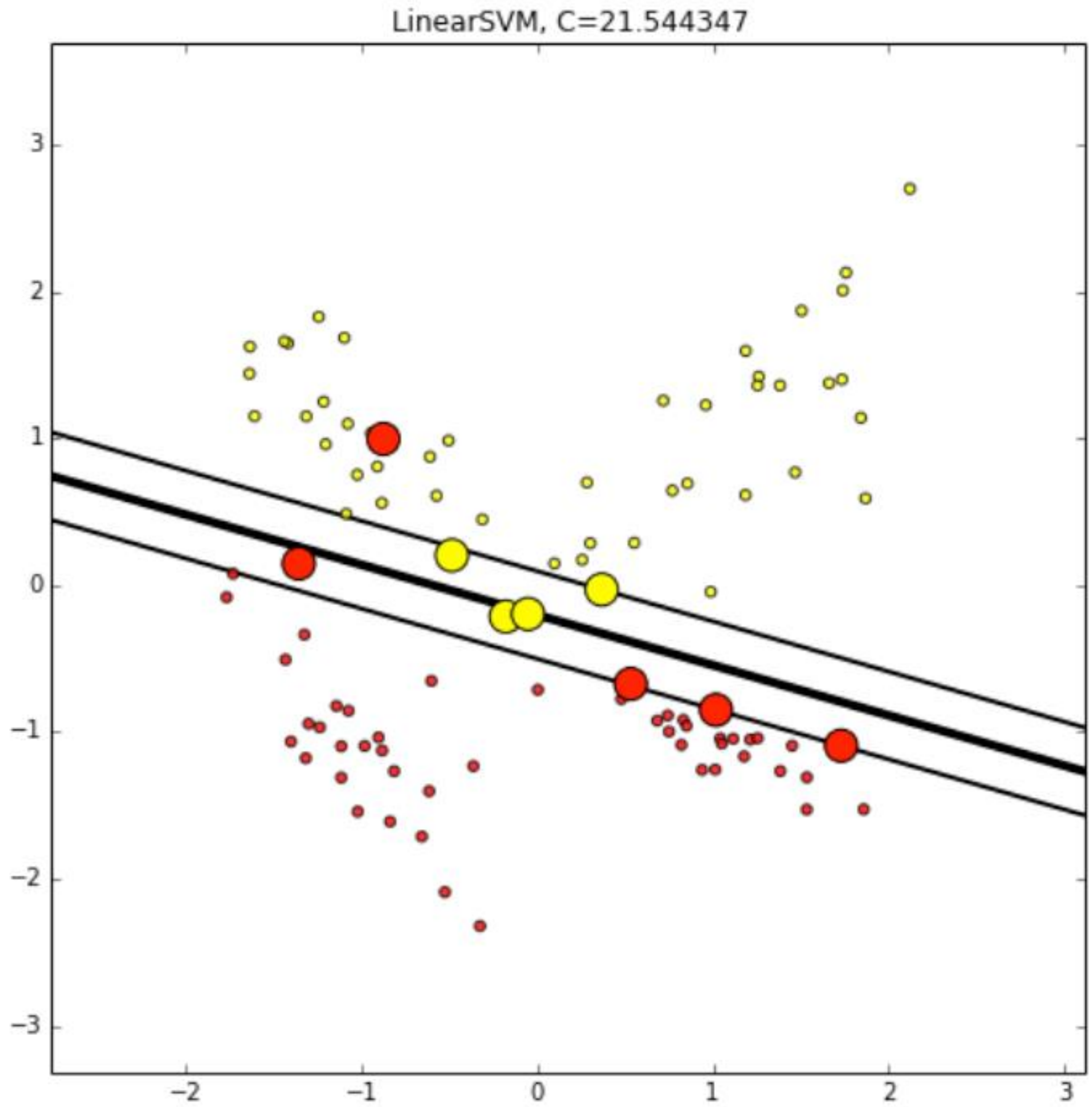
ЗНАЧЕНИЕ КОНСТАНТЫ С



ЗНАЧЕНИЕ КОНСТАНТЫ C



ЗНАЧЕНИЕ КОНСТАНТЫ C



ТИПЫ ОБЪЕКТОВ В SVM