

Лекция 6

Метод опорных векторов

Дубенюк Анна Антоновна
anya.dubenyuk@yandex.ru
@andu192

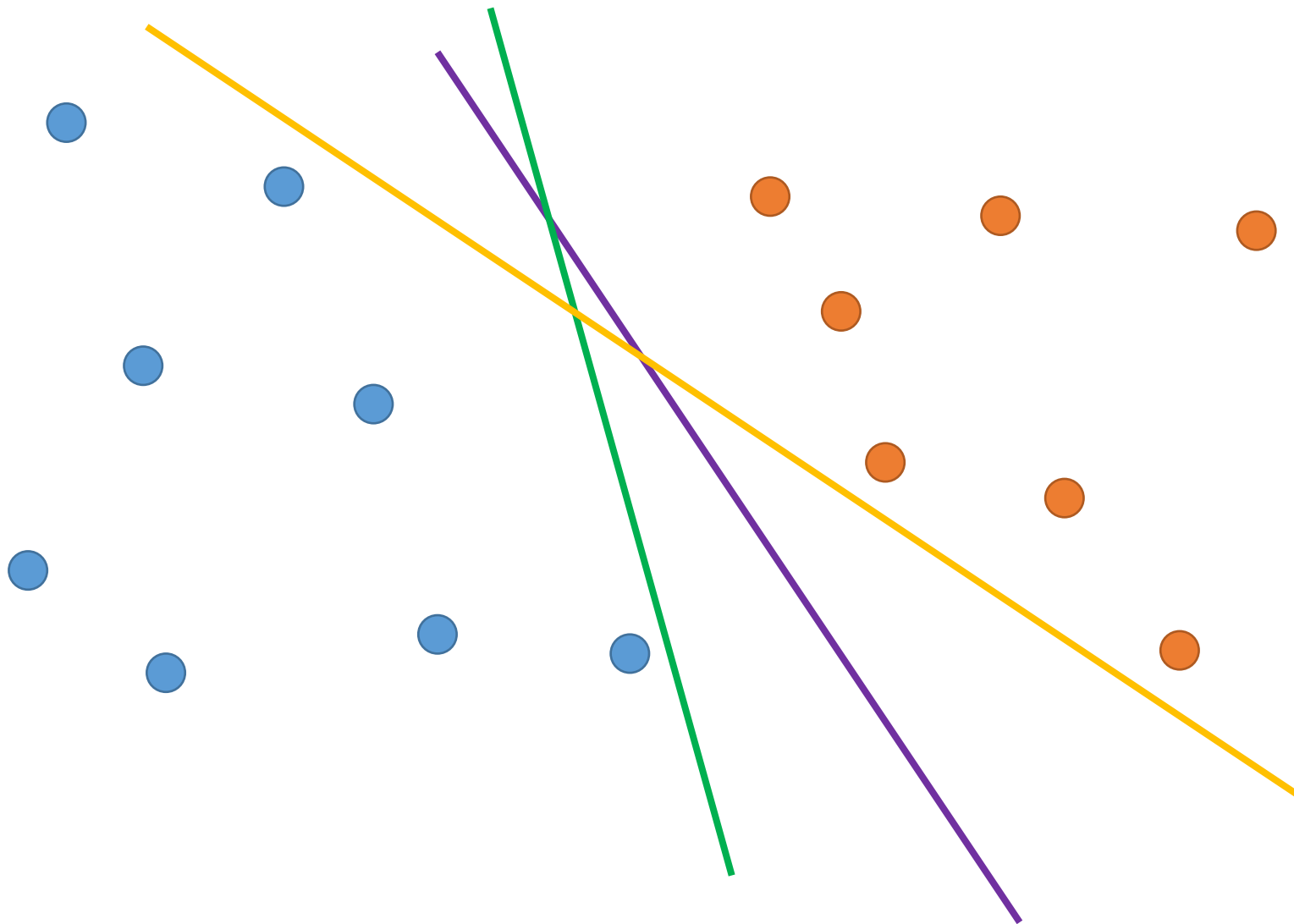
ВШЭ, 2023

Hinge loss

- Решаем задачу бинарной классификации: $\mathbb{Y} = \{-1, +1\}$
- Минимизация верхней оценки:

$$\frac{1}{\ell} \sum_{i=1}^{\ell} \max(0, 1 - y_i \langle w, x_i \rangle) \rightarrow \min_w$$

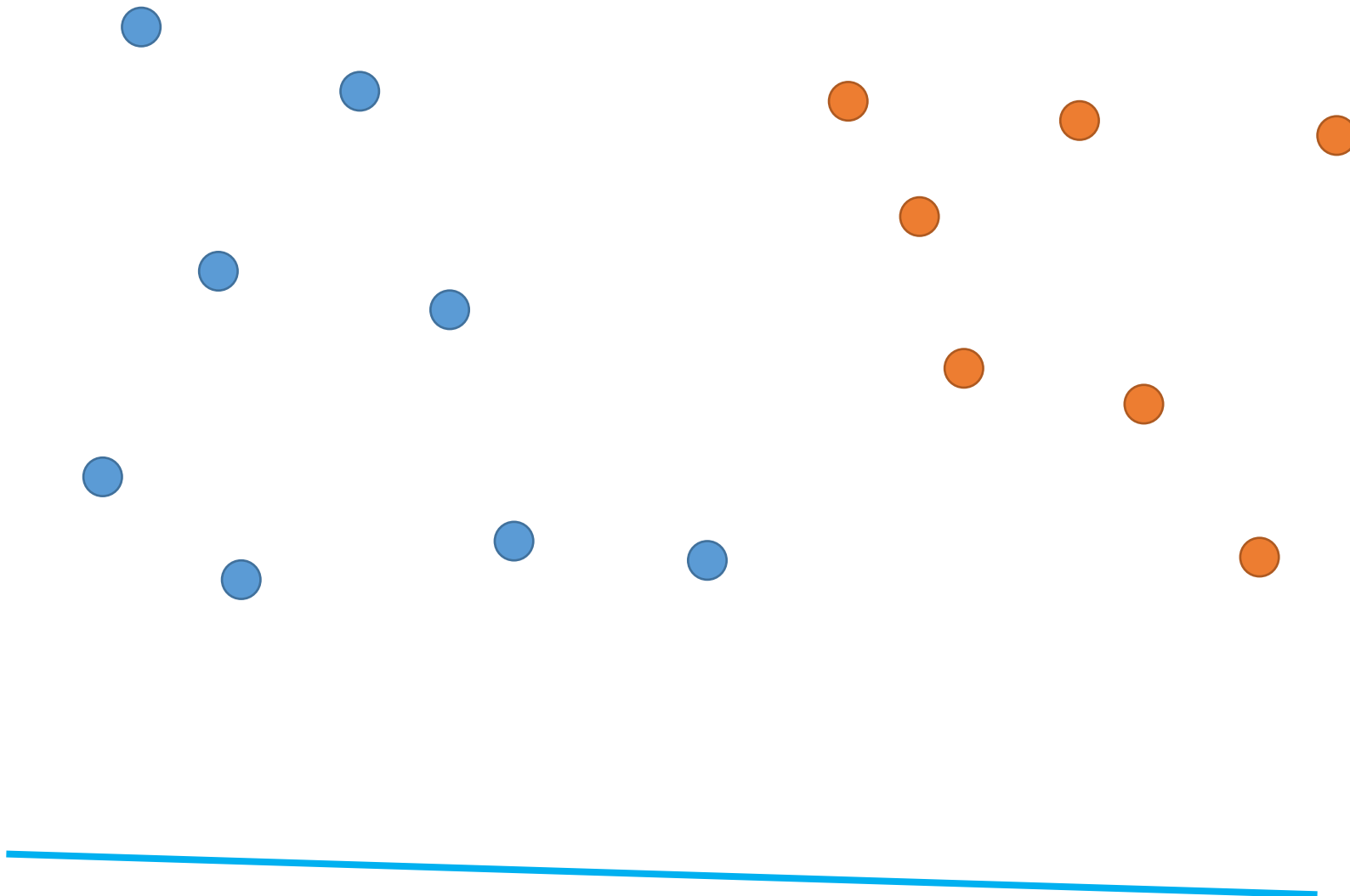
Какой классификатор лучше?



Отступ классификатора

- Будем максимизировать отступ классификатора — расстояние от гиперплоскости до ближайшего объекта

Отступ классификатора



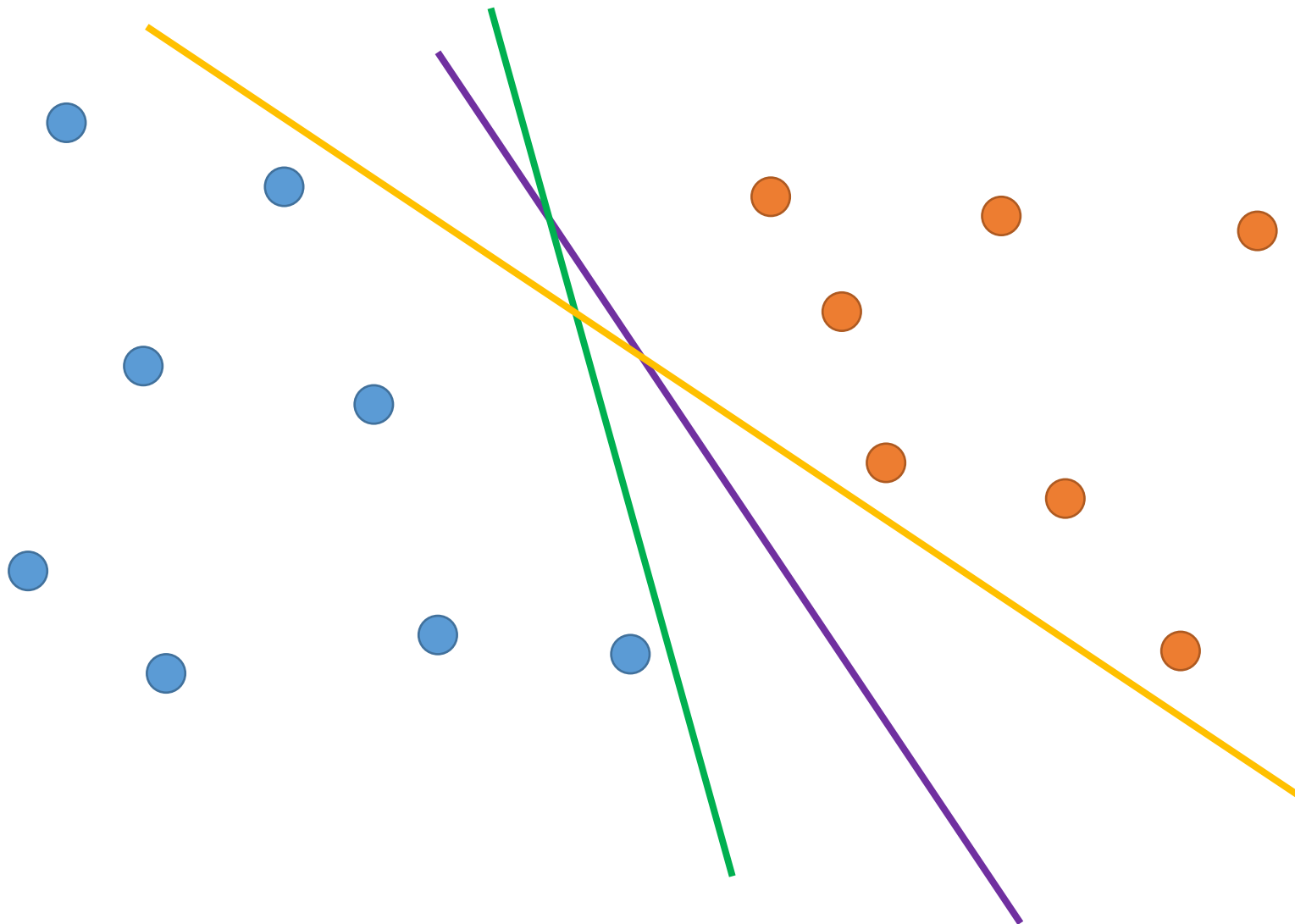
Отступ классификатора

- Будем максимизировать отступ классификатора — расстояние от гиперплоскости до ближайшего объекта
- При этом будет стараться сделать поменьше ошибок
- По сути, делаем как можно меньше предположений о модели, и верим, что это понизит вероятность переобучения

Простой случай

- Будем считать, что выборка линейно разделима
- Существует линейный классификатор, не допускающий ни одной ошибки

Линейно разделимый случай



Линейно разделимый случай

- **Требование 1:** $y_i(\langle w, x_i \rangle + w_0) > 0$ для всех $i = 1, \dots, \ell$
- **Требование 2:** максимальный отступ классификатора

Отступ классификатора

- Расстояние от точки до гиперплоскости $\langle w, x \rangle + w_0 = 0$:

$$\frac{|\langle w, x \rangle + w_0|}{\|w\|}$$

- Отступ классификатора:

$$\min_{i=1, \dots, \ell} \frac{|\langle w, x_i \rangle + w_0|}{\|w\|}$$

Небольшое предположение

- Линейный классификатор:

$$a(x) = \text{sign} (\langle w, x_i \rangle + w_0)$$

- Если мы поделим w и w_0 на число $a > 0$, то выходы классификатора никак не поменяются:

$$a(x) = \text{sign} \left(\frac{\langle w, x_i \rangle + w_0}{a} \right) = \text{sign} (\langle w, x_i \rangle + w_0)$$

Небольшое предположение

- Поделим w и w_0 на $\min_{i=1,\dots,\ell} |\langle w, x_i \rangle + w_0| > 0$, после этого будет выполнено

$$\min_{i=1,\dots,\ell} |\langle w, x_i \rangle + w_0| = 1$$

Отступ классификатора

- Расстояние от точки до гиперплоскости $\langle w, x \rangle + w_0 = 0$:

$$\frac{|\langle w, x \rangle + w_0|}{\|w\|}$$

- Отступ классификатора:

$$\min_{i=1, \dots, \ell} \frac{|\langle w, x_i \rangle + w_0|}{\|w\|} = \frac{\min_{i=1, \dots, \ell} |\langle w, x_i \rangle + w_0|}{\|w\|} = \frac{1}{\|w\|}$$

Линейно разделимый случай

- **Требование 1:** $y_i(\langle w, x_i \rangle + w_0) > 0$ для всех $i = 1, \dots, \ell$
- **Требование 2:** максимальный отступ классификатора

$$\frac{1}{\|w\|} \rightarrow \max_w$$

Линейно разделимый случай

- **Требование 1:** $y_i(\langle w, x_i \rangle + w_0) > 0$ для всех $i = 1, \dots, \ell$
- **Требование 2:** максимальный отступ классификатора

$$\frac{1}{\|w\|} \rightarrow \max_w$$

- При условии, что $\min_{i=1, \dots, \ell} |\langle w, x_i \rangle + w_0| = 1$

Линейно разделимый случай

- **Требование 1:** $y_i(\langle w, x_i \rangle + w_0) > 0$ для всех $i = 1, \dots, \ell$
- **Требование 2:** максимальный отступ классификатора

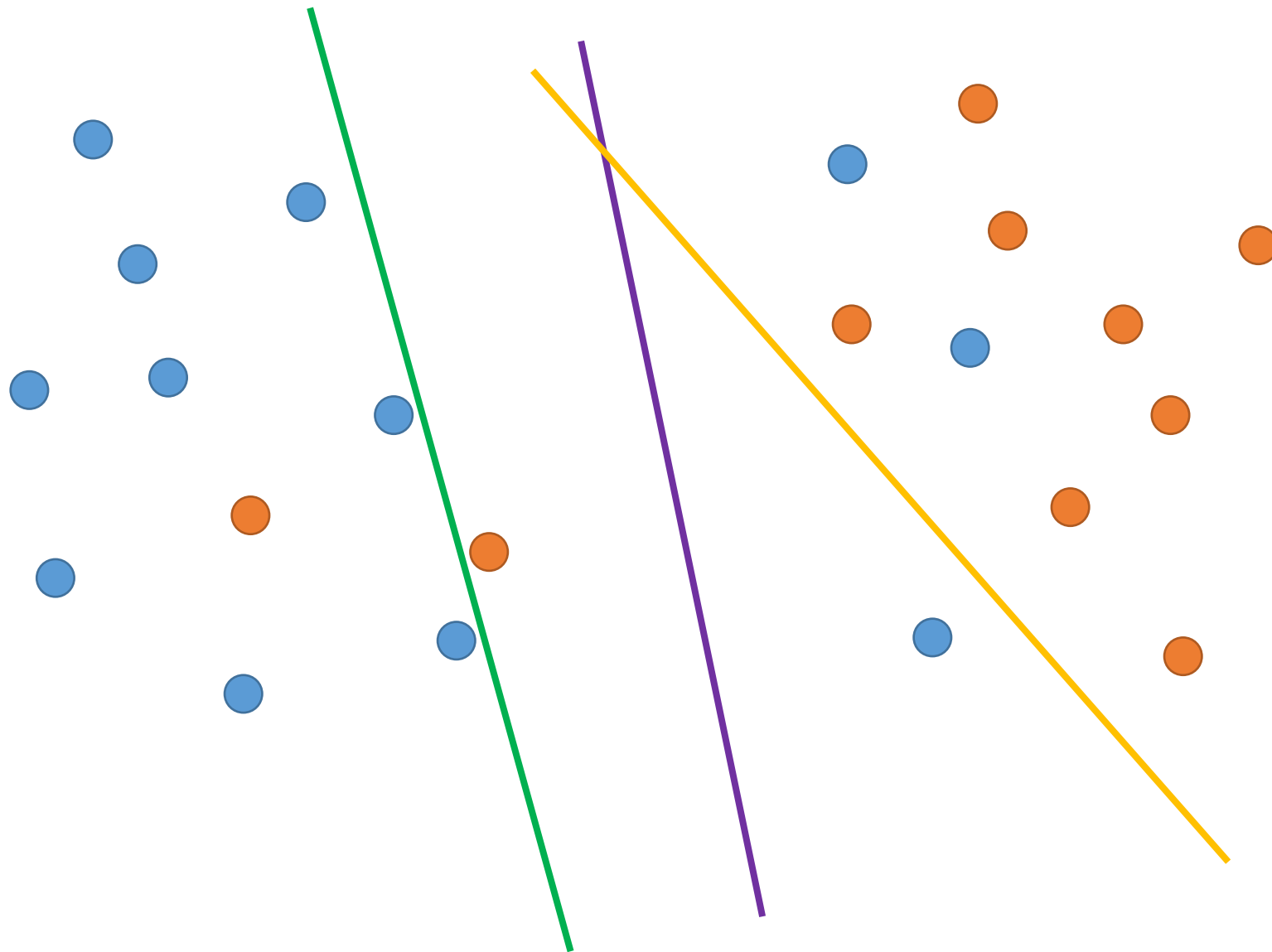
$$\frac{1}{\|w\|} \rightarrow \max_w$$

- При условии, что $|\langle w, x_i \rangle + w_0| \geq 1$
- И мы минимизируем $\|w\|$ — тогда где-то модуль отступа будет равен 1

Метод опорных векторов (SVM)

$$\begin{cases} \|w\|^2 \rightarrow \min_{w, w_0} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 \end{cases}$$

Линейно неразделимый случай



Линейно неразделимый случай

- Любой линейный классификатор допускает хотя бы одну ошибку

$$\begin{cases} \|w\|^2 \rightarrow \min_{w, w_0} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 \end{cases}$$

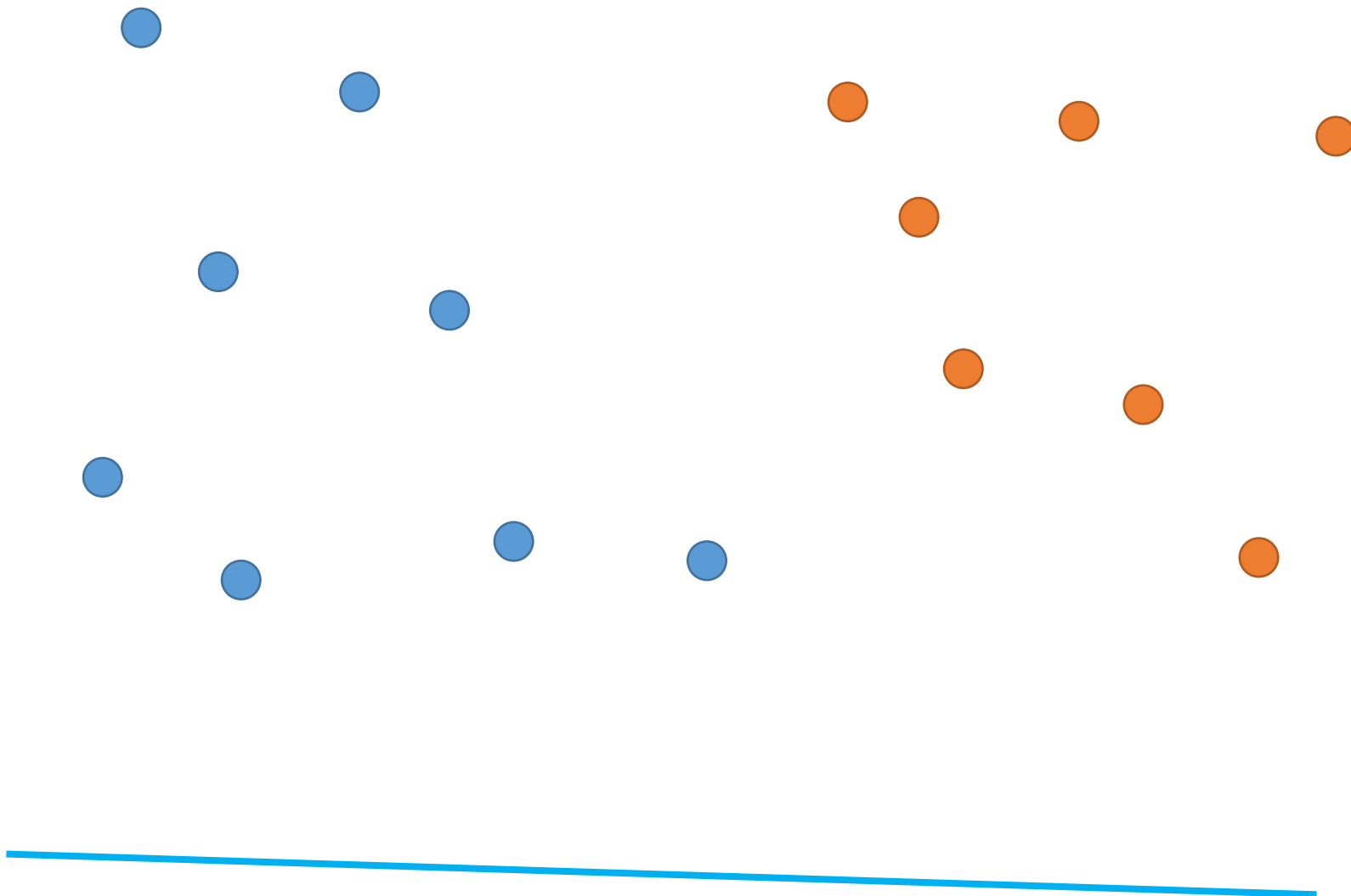
Линейно неразделимый случай

$$\begin{cases} \|w\|^2 \rightarrow \min_{w, w_0} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

Линейно неразделимый случай

$$\begin{cases} \|w\|^2 \rightarrow \min_{w, w_0} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 - 10^{1000} \end{cases}$$

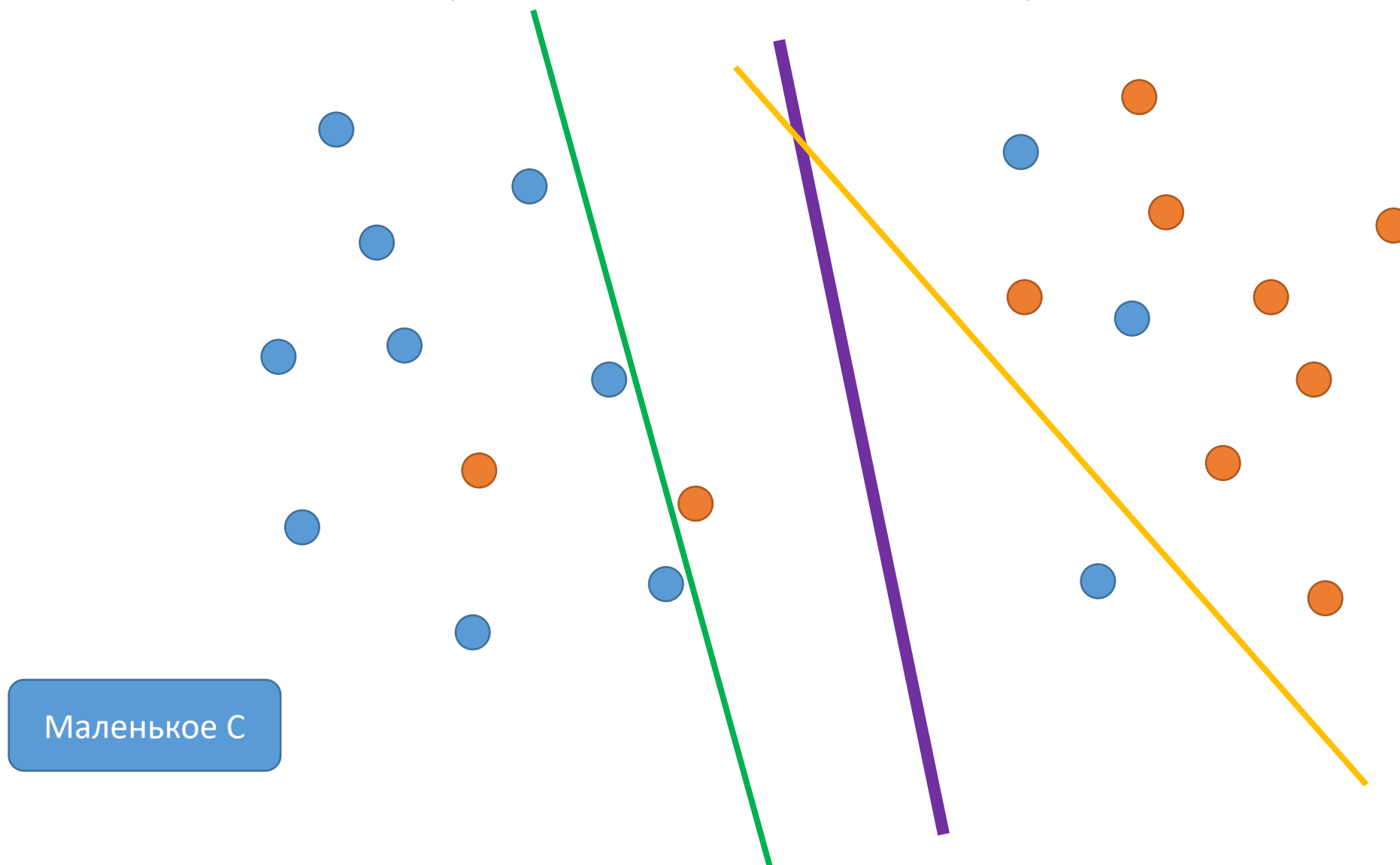
Отступ классификатора



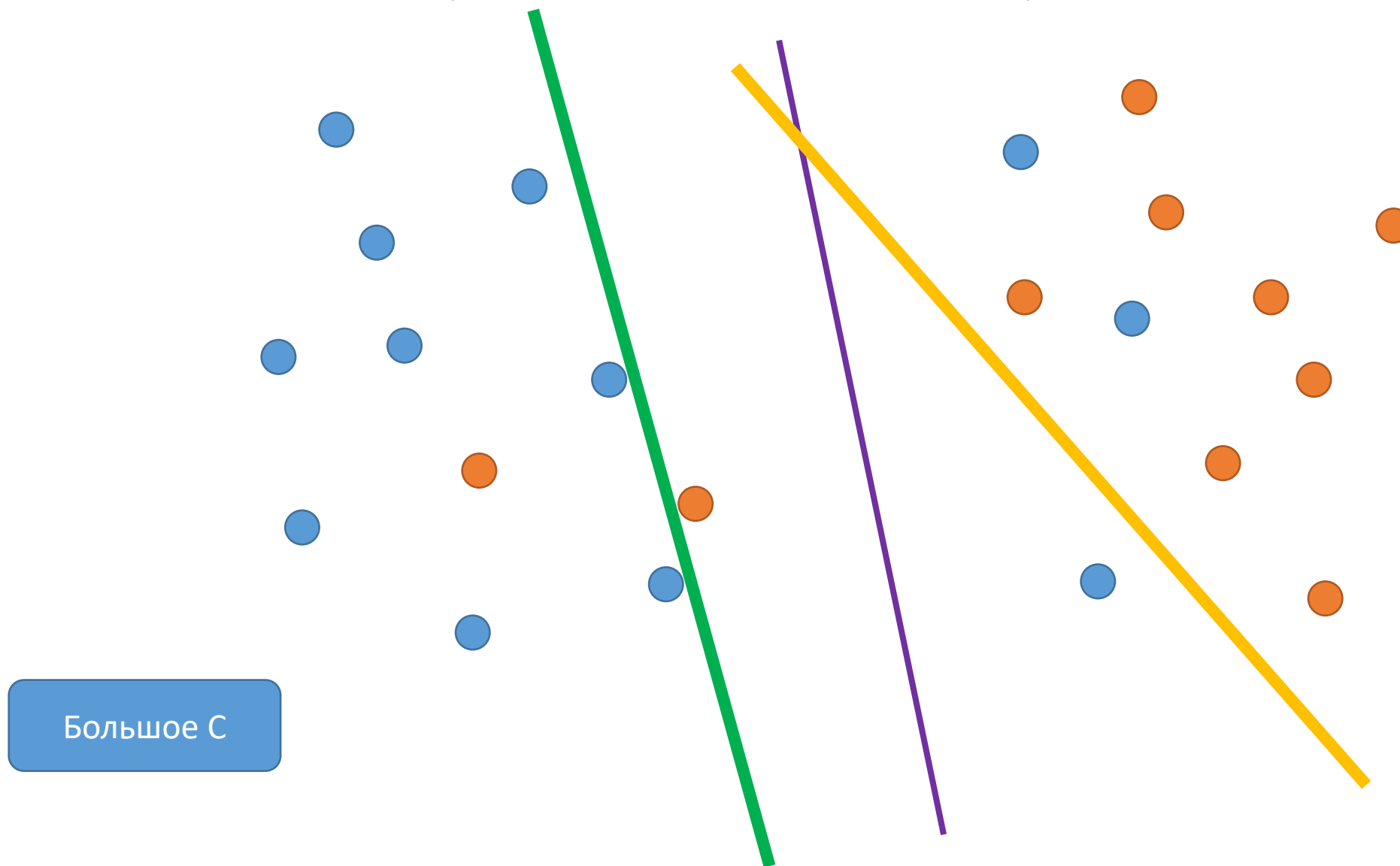
Метод опорных векторов

$$\left\{ \begin{array}{l} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, w_0, \xi_i} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{array} \right.$$

Линейно неразделимый случай



Линейно неразделимый случай



Метод опорных векторов

$$\begin{cases} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, w_0, \xi_i} \\ y_i(\langle w, x_i \rangle + w_0) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

- Объединим ограничения:

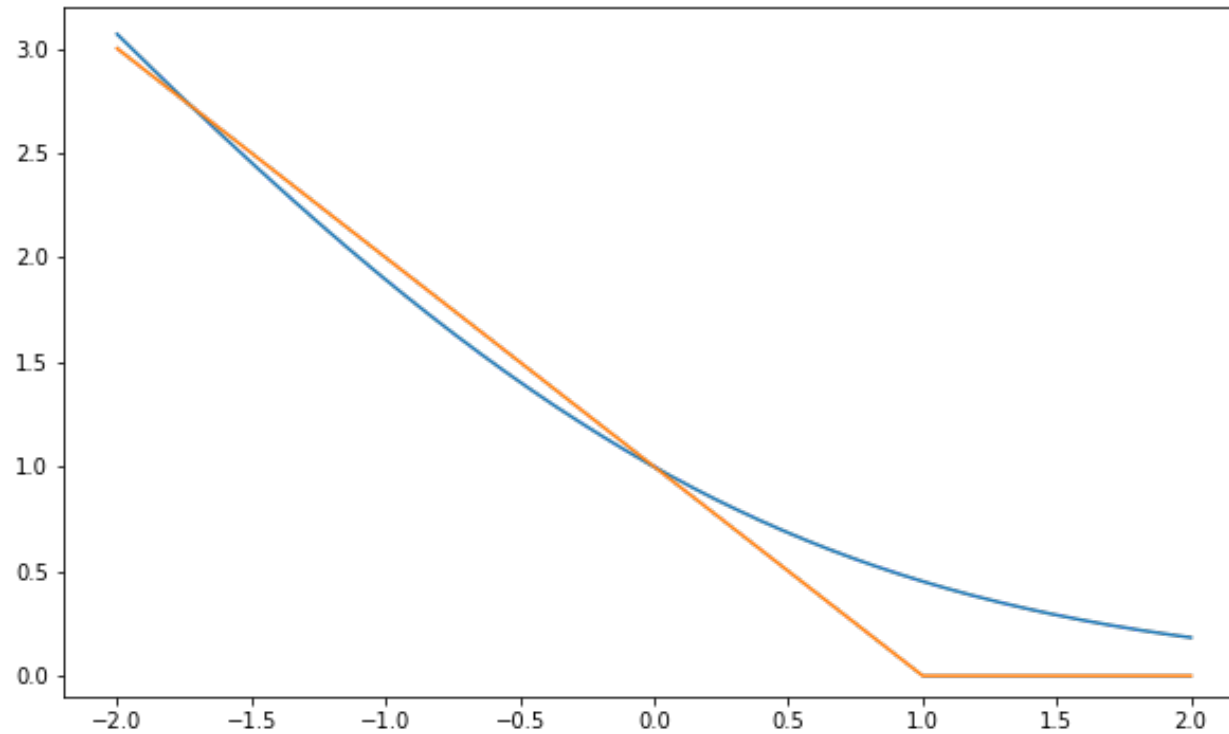
$$\xi_i \geq \max(0, 1 - y_i(\langle w, x_i \rangle + w_0))$$

Метод опорных векторов

$$C \sum_{i=1}^{\ell} \max(0, 1 - y_i(\langle w, x_i \rangle + w_0)) + \|w\|^2 \rightarrow \min_{w, w_0}$$

- Функция потерь (hinge loss) + регуляризация

Сравнение логистической регрессии и SVM



Резюме

- Логистическая регрессия — обучение модели так, что на объектах с близкими прогнозами эти прогнозы стремятся к доле положительных объектов
- Метод опорных векторов основан на идее максимизации отступа классификатора