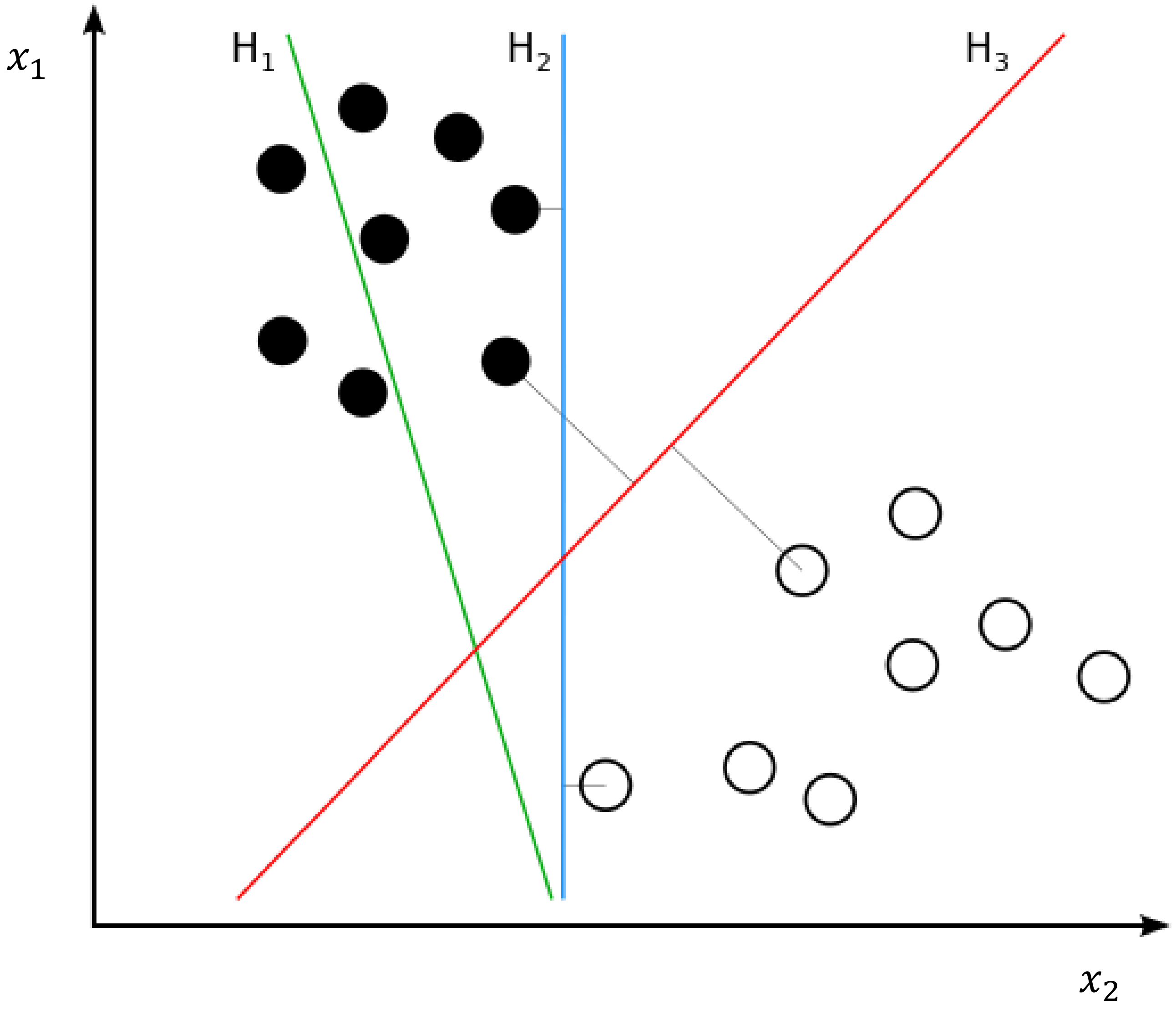
SVM

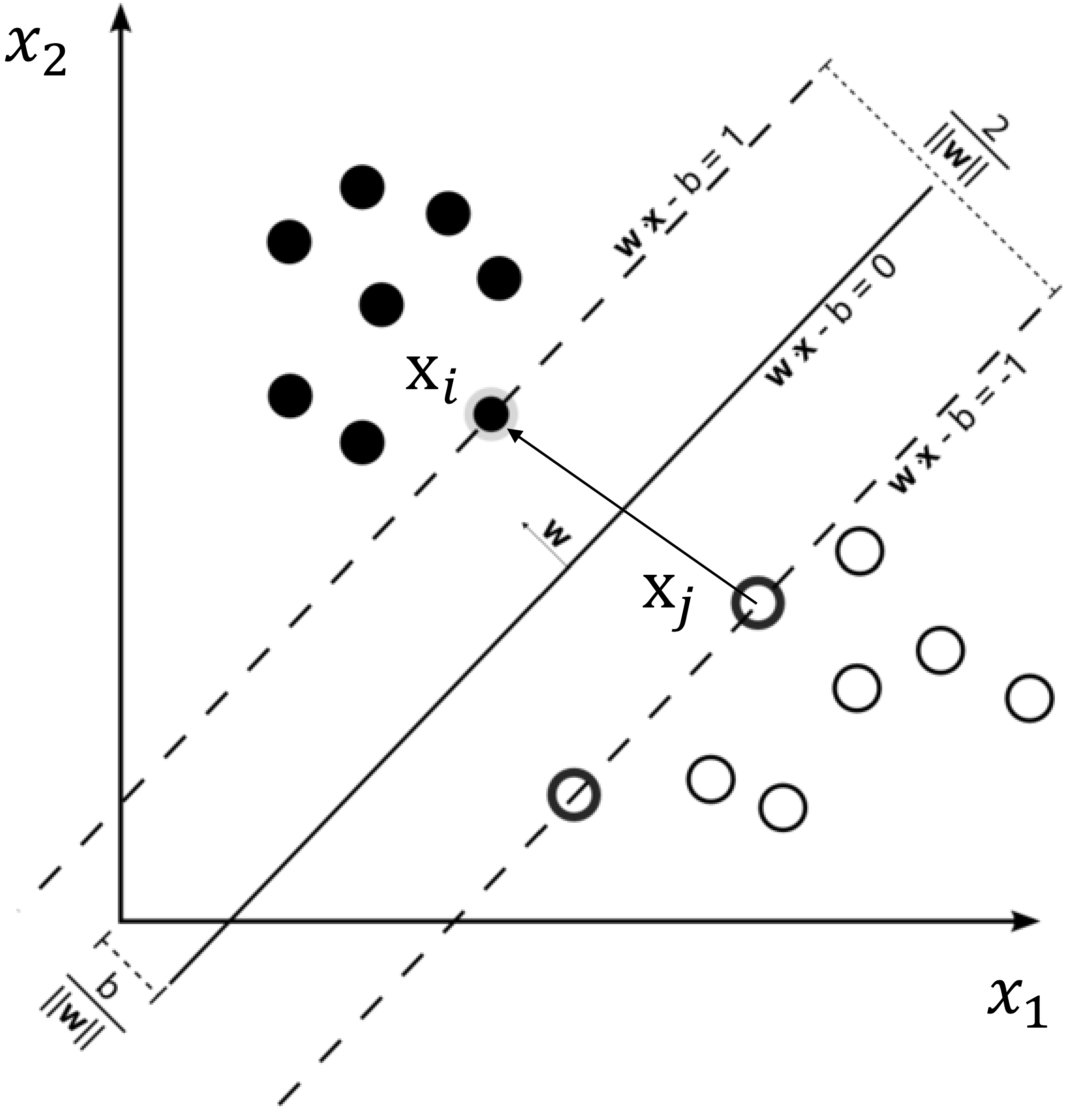
Support Vector Machines

Метод Опорных Векторов

SVM (Support Vector Machines)



Максимизация отступа
Линейно разделимая выборка



Нормализуем **w** и ***b***, так чтобы,

$$\min\left|\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{b}\right| = \min\left(\boldsymbol{y}_i\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{b}\right)\right) = \mathbf{1},$$

Тогда максимальная ширина полосы – проекция вектора $\mathbf{x}_i - \mathbf{x}_j$ на нормаль $\frac{\mathbf{w}}{||\mathbf{w}||}$, т.е:

$$\frac{\mathbf{w}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{x}_j)}{||\mathbf{w}||} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{b} - (\mathbf{w}^{\mathrm{T}}\mathbf{x}_2 - \boldsymbol{b})}{||\mathbf{w}||} = \frac{\mathbf{2}}{||\mathbf{w}||}$$

Тогда задача – это максимизация $\frac{2}{||\boldsymbol{w}||}$

или минимизация $\left||\boldsymbol{w}|\right|$, или $\mathbf{w}^{\mathrm{T}}\mathbf{w}$

при условии $\boldsymbol{y}_i(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - \boldsymbol{b}) \geq \mathbf{1}$

Задача оптимизации

$$\begin{cases} \dfrac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) \geq 1 \end{cases}$$

Теорема Каруша-Куна-Такера (ККТ)

Задача оптимизации с ограничениями:

$$\begin{cases} \min\limits_{z} f(z) \\ g_i(z) \le 0 \\ h_i(z) = 0 \end{cases}$$

Если z^* – точка локального минимума, то существуют множители Лагранджа α_i^* и β_j^* для:

$$\mathcal{L}(z, \alpha, \beta) = f(z) + \sum_{i=1}^{m} \alpha_i g_i(z) + \sum_{j=1}^{k} \beta_j h_j(z),$$

такие, что :

$$\begin{cases} \frac{\partial}{\partial z_i} \mathcal{L}(z^*, \alpha^*, \beta^*) = 0, \\ \frac{\partial}{\partial \beta_i} \mathcal{L}(z^*, \alpha^*, \beta^*) = 0, \\ \alpha_i g_i(z^*) = 0, \\ \alpha_i^* \ge 0 \end{cases}$$

И задачу можно свести к двойственной: $\max\limits_{\alpha,\beta} \mathcal{L}(z, \alpha, \beta)$

Двойственная задача оптимизации

$$\begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) \geq 1 \end{cases} \to \begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ -\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) \leq 0 \end{cases}$$

$$\Big\downarrow$$

$$\mathcal{L}(\overbrace{\mathrm{w}, b}^{\mathrm{z}}, \alpha) = \frac{1}{2}\left(\mathrm{w}^{\mathrm{T}}\mathrm{w}\right) - \sum_{i=1}^{N} \alpha_i \left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right);$$

Двойственная задача оптимизации

$$\begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) \geq 1 \end{cases} \to \begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ -\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) \leq 0 \end{cases}$$

$$\downarrow$$

$$\mathcal{L}(\overbrace{\mathrm{w}, b}^{\mathrm{z}}, \alpha) = \frac{1}{2}\left(\mathrm{w}^{\mathrm{T}}\mathrm{w}\right) - \sum_{i=1}^{N} \alpha_i \left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right);$$

$$\alpha_i \geq 0; \qquad \alpha_i\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) = 0$$

Решение двойственной задачи

$$\mathcal{L}(\mathrm{w}, \mathrm{\alpha}, b) = \frac{1}{2}\left(\mathrm{w}^{\mathrm{T}}\mathrm{w}\right) - \sum_{i=1}^{N} \alpha_i \left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right); \qquad \alpha_i \geq 0; \qquad \alpha_i\left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) = 0$$

$$\nabla_{\mathrm{w}}\mathcal{L}(\mathrm{w}, \mathrm{\alpha}, b) = \mathrm{w} - \sum_{i=1}^{N} \alpha_i\, y_i \mathrm{x}_i, = 0 \;\; \rightarrow \;\; \mathrm{w} = \sum_{i=1}^{N} \alpha_i\, y_i \mathrm{x}_i$$

$$\frac{\partial}{\partial b}\mathcal{L}(\mathrm{w}, \mathrm{\alpha}, b) = \sum_{i=1}^{N} \alpha_i\, y_i = 0$$

$$\mathcal{L}(\mathrm{w}, \mathrm{\alpha}, b) = \frac{1}{2}\left(\mathrm{w}^{\mathrm{T}}\mathrm{w}\right) - \sum_{i=1}^{N} \alpha_i \left(y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) = \frac{1}{2}\sum\sum y_i y_j \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j - \sum\sum y_i y_j \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j + \sum \alpha_i =$$

$$= \mathcal{L}(\mathrm{w}, \mathrm{\alpha}, b) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} y_i y_j\, \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j$$

Алгоритм оптимизации – выбираем пару α_i, α_j и оптимизируем $\mathcal{L}$ при остальных α фиксированных.

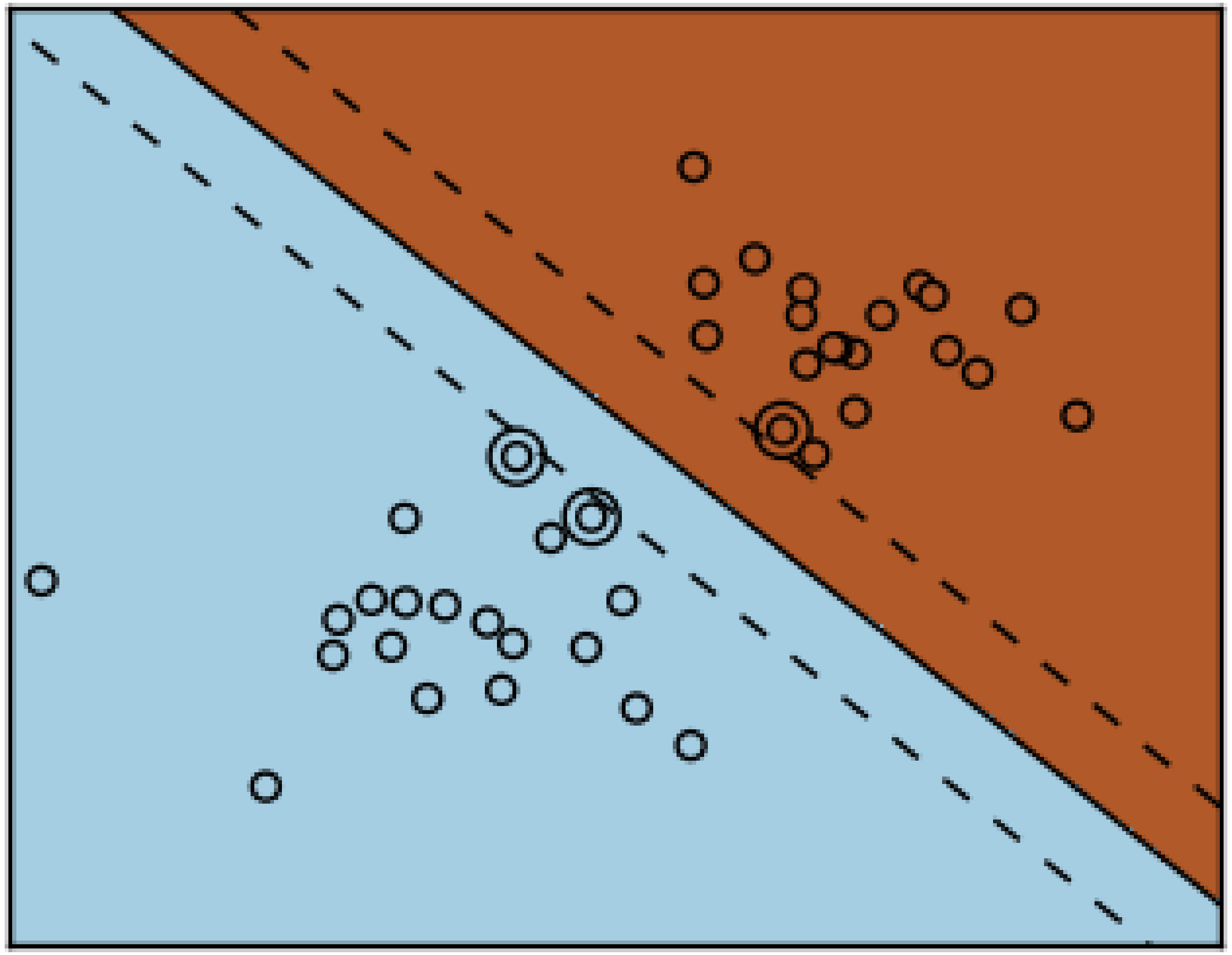
Опорные вектора

$$\mathrm{w} = \sum_{i=1}^{N} \alpha_i \, y_i \mathrm{x}_i$$

$$\alpha_i \big(y_i (\mathrm{w}^{\mathrm{T}} \mathrm{x}_i - b) - 1 \big) = 0$$

$$\mathrm{x}_i \colon \alpha_i > 0$$

$$y_i (\mathrm{w}^{\mathrm{T}} \mathrm{x}_i - b) - 1 = 0$$



Неразделимая выборка

Линейно разделимая выборка:

$$\begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} \to \min \\ y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) \geq 1 \end{cases}$$

Неразделимая выборка:

$$\begin{cases} \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} + C\sum_{i=1}^{N}\xi_i \to \min \\ y_i\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}$$

Применение ККТ к задаче SVM на неразделимой выборке

$$\mathcal{L}(\overbrace{\mathrm{w}, b, \xi}^{\mathrm{z}}, \underbrace{\alpha, r}_{\alpha}) = \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i(y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) - 1 + \xi_i) - \sum_{j=1}^{N} r_j\xi_i$$

$$= \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} - \sum \alpha_i(y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) - 1) - \sum \xi_i(r_i + \alpha_i - C)$$

$$\alpha_i, r_i \geq 0; \qquad \alpha_i\,(y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) - 1 + \xi_i) = 0; \qquad r_i\,\xi_i = 0$$

$$\frac{\partial\mathcal{L}}{\partial\mathrm{w}} = \frac{\partial\mathcal{L}}{\partial b} = \frac{\partial\mathcal{L}}{\partial\xi} = 0 \Rightarrow \begin{cases} \mathrm{w} = \sum \alpha_i\,\mathrm{y}_i\mathrm{x}_i \\ \sum \alpha_i y_i = 0 \\ \alpha_i = C - r_i \Rightarrow 0 \leq \alpha_i \leq C \end{cases}$$

Вид двойственной задачи

$$\begin{cases} \mathrm{w} = \sum \alpha_i \, \mathrm{y}_i \mathrm{x}_i \\ \sum \alpha_i y_i = 0 \\ \alpha_i = C - r_i \end{cases} \Rightarrow$$

$$\mathcal{L}(\alpha) = \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} - \sum \alpha_i \left(y_i \left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b\right) - 1\right) - \sum \xi_i (r_i + \alpha_i - C) =$$

$$= \frac{1}{2}\sum\sum \mathrm{y}_i \mathrm{y}_j \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}} \mathrm{x}_j - \sum\sum \mathrm{y}_i \mathrm{y}_j \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}} \mathrm{x}_j + \sum \alpha_i$$

$$\begin{cases} \max\limits_{\alpha} \mathcal{L}(\alpha) = \sum \alpha_i - \frac{1}{2}\sum\sum \mathrm{y}_i \mathrm{y}_j \alpha_i \alpha_j \mathrm{x}_i^{\mathrm{T}} \mathrm{x}_j \\ 0 \le \alpha_i \le C \\ \sum \alpha_i y_i = 0 \end{cases}$$

Типы объектов

$$\mathrm{w} = \sum \alpha_i \mathrm{y}_i \mathrm{x}_i; \qquad y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) \geq 1 - \xi_i$$

$$\alpha_i = C - r_i; \qquad \alpha_i\,(y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) - 1 + \xi_i) = 0; \qquad r_i \xi_i = 0$$

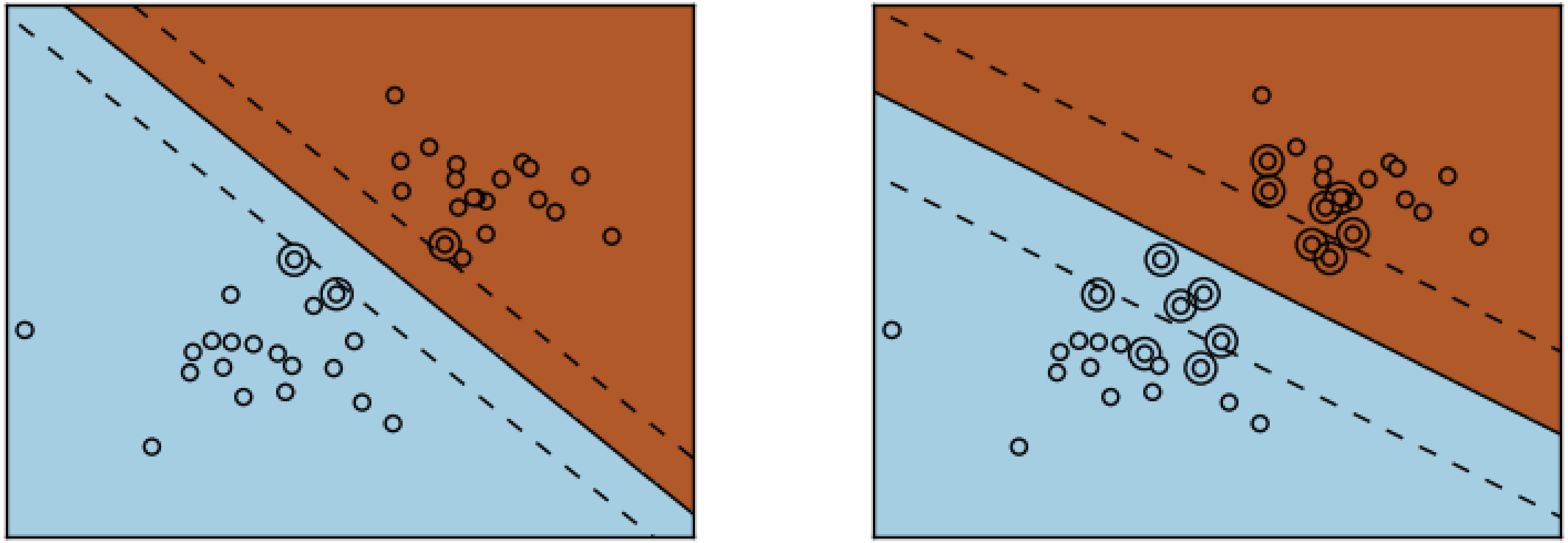
1. $\alpha_i = 0;\ \xi_i = 0;\ y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) \geq 1$ – Внутренние объекты

2. $0 < \alpha_i < C;\ \xi_i = 0;\ y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) = 1$ – Опорные объекты

3. $\alpha_i = C;\ \xi_i > 0;\ y_i(\mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b) \leq 1$ – Опорные объекты – нарушители

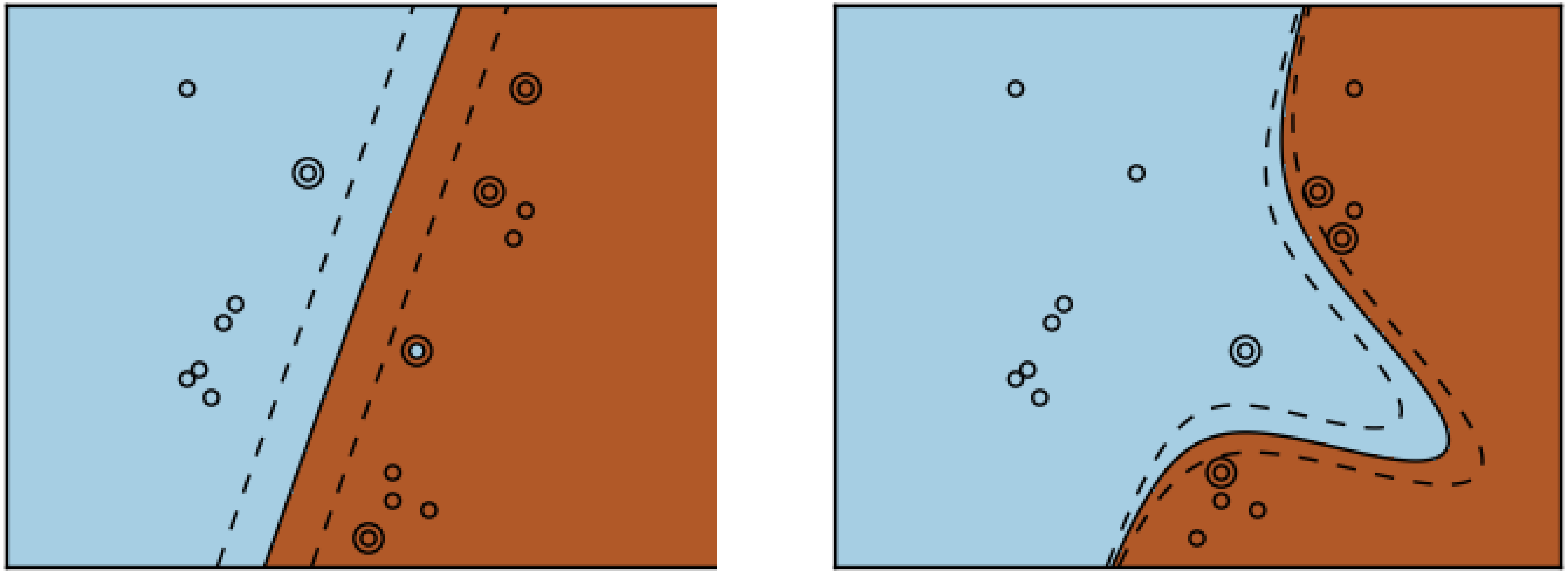
Влияние константы

$$\min\left(\frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} + C\sum_{i=1}^{N}\xi\right)$$



$C \uparrow$ $C \downarrow$

Более высокая размерность и ошибка



$$E_{gen} \leq \frac{\#\ of\ support\ vectors}{\#\ of\ training\ vectors}$$

Kernel trick (Ядерный трюк)

Заметим, что во всех расчетах мы используем только $\mathrm{x}_i^{\mathrm{T}}\mathrm{x}_j$, тогда, для перехода к более высокой размерности, нужно лишь определить скалярное произведение векторов.

$K(\mathrm{x}, \mathrm{x}') - \text{ядро, если } K(\mathrm{x}, \mathrm{x}') = \psi(\mathrm{x})^{\mathrm{T}}\psi(\mathrm{x}'), \text{при } \psi\colon \mathrm{X} \to \mathrm{H}, \text{где H} - \text{гильбертово пространство.}$

Пример:

$$K(\mathrm{x}, \mathrm{x}') = \left(1 + \mathrm{x}^{\mathrm{T}}\mathrm{x}'\right)^2 = 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x'_1 + 2x_2 x'_2 + 2x_1 x'_1 x_2 x'_2$$

$$\psi(\mathrm{x}) = (1, x_1^2, x_2^2, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2)$$

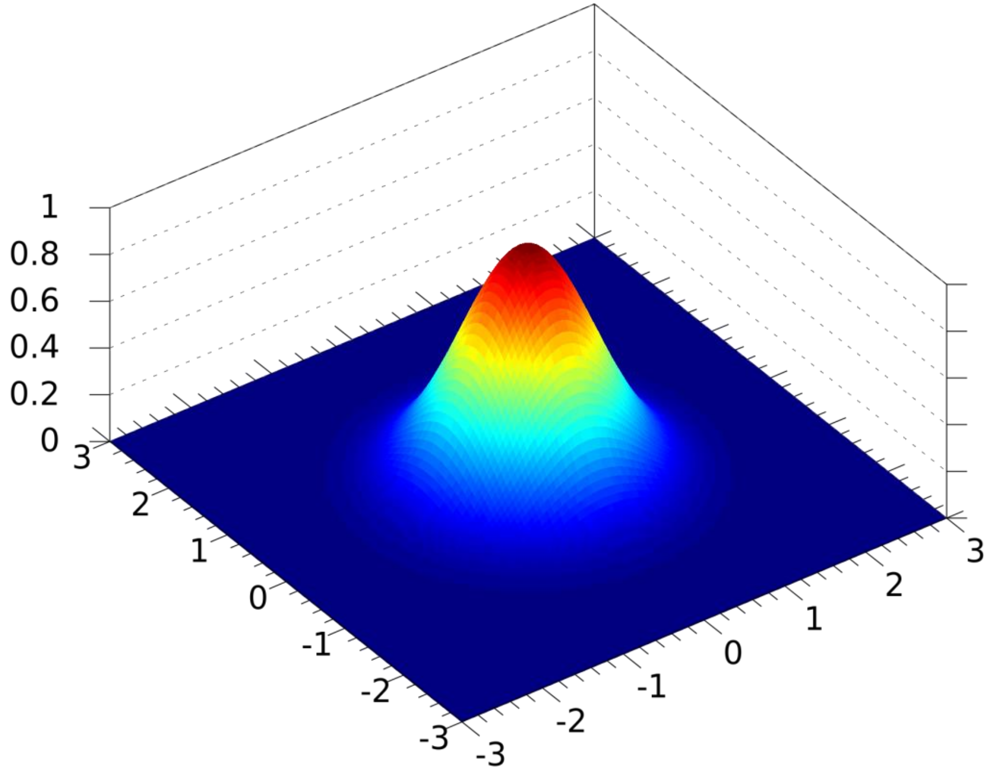
Разные ядра

Линейное ядро: $\langle \mathrm{x}, \mathrm{x}' \rangle$

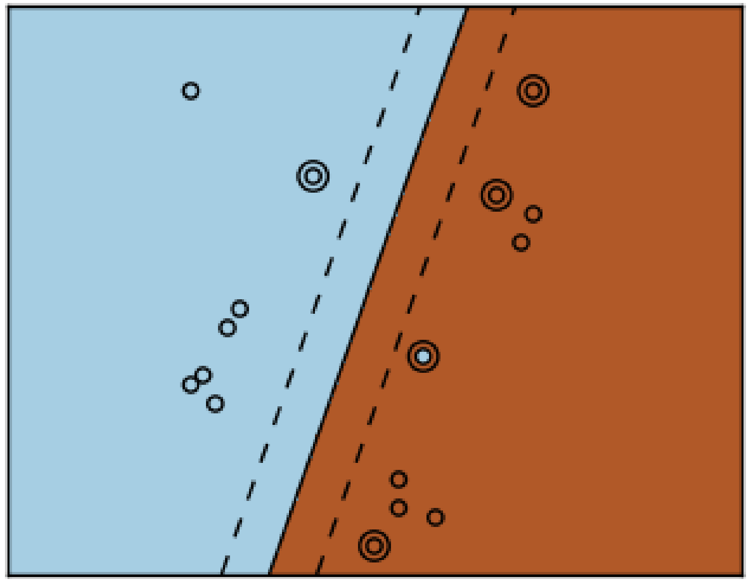
Полиномиальное ядро: $(r + \gamma \langle \mathrm{x}, \mathrm{x}' \rangle)^d$

Гауссианово ядро (RBF): $e^{-\gamma |\mathrm{x}-\mathrm{x}'|^2}$

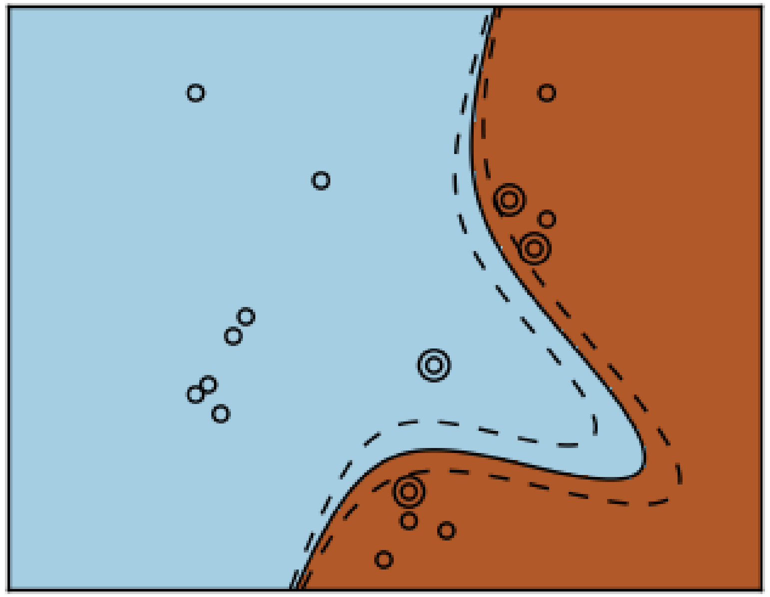
Сигмоид: $\mathrm{th}(\gamma \langle \mathrm{x}, \mathrm{x}' \rangle + r)$



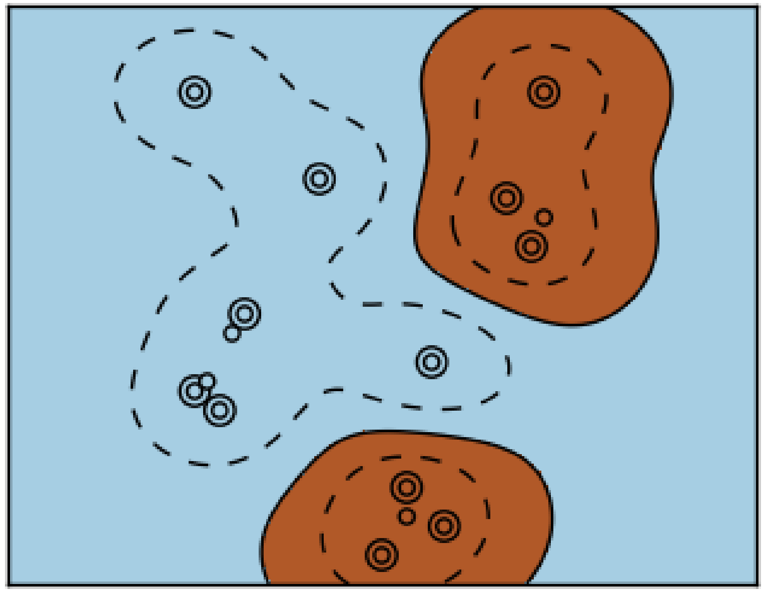
Radial Basis Function (RBF)



Линейное ядро



Полиномиальное ядро



Гауссианово ядро (RBF)

Support Vector Networks

Cortes and Vapnik, 1995