# Введение в искусственный интеллект. Машинное обучение

## Тема семинара: пример расчетов для SVM

Бабин Д.Н., Иванов И.Е., Петюшко А.А.

кафедра Математической Теории Интеллектуальных Систем

# План семинара

1. Постановка задачи

# План семинара

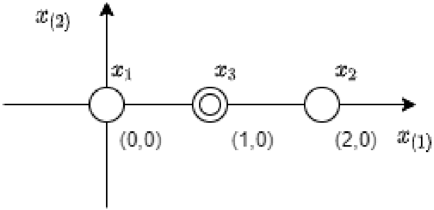
1. Постановка задачи
2. Выбор нелинейного ядра

# План семинара

1. Постановка задачи
2. Выбор нелинейного ядра
3. Необходимые расчеты

## Задача

Методом опорных векторов разделить классы $A = \{x_1, x_2\}$ и $B = \{x_3\}$, если $x_1 = (0, 0), x_2 = (2, 0), x_3 = (1, 0)$. Полагаем, что $y_1 = y_2 = +1, y_3 = -1$.



# Выбор подходящего ядра

## Указание [доказать]

Любые $m+1$ точек из $\mathbb{R}^n$ могут быть линейно разделены на любые два класса после перехода в новое пространство с помощью мономиального отображения степени не больше $m$ (и соответственно ядра $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^m$).

# Выбор подходящего ядра

**Указание [доказать]**

Любые $m+1$ точек из $\mathbb{R}^n$ могут быть линейно разделены на любые два класса после перехода в новое пространство с помощью мономиального отображения степени не больше $m$ (и соответственно ядра $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^m$).

Поэтому отображение ищем в виде
$\varphi : X \to Y, Y = \mathbb{R}^k, z \in Y, z_{(i)} = x_{(1)}^{i_1} x_{(2)}^{i_2} \dots x_{(n)}^{i_n}|_{i_1+i_2+\dots+i_n \leq m}, 1 \leq i \leq k$ для
$X = \mathbb{R}^n, n = 2, m = 2.$

# Выбор подходящего ядра

**Указание [доказать]**

Любые $m+1$ точек из $\mathbb{R}^n$ могут быть линейно разделены на любые два класса после перехода в новое пространство с помощью мономиального отображения степени не больше $m$ (и соответственно ядра $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^m$).

Поэтому отображение ищем в виде
$\varphi : X \to Y, Y = \mathbb{R}^k, z \in Y, z_{(i)} = x_{(1)}^{i_1} x_{(2)}^{i_2} \dots x_{(n)}^{i_n}|_{i_1+i_2+\dots+i_n \le m}, 1 \le i \le k$ для
$X = \mathbb{R}^n, n = 2, m = 2.$
Ядро, соответствующее этому отображению: $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^2$.

# Выбор подходящего ядра

## Указание [доказать]

Любые $m+1$ точек из $\mathbb{R}^n$ могут быть линейно разделены на любые два класса после перехода в новое пространство с помощью мономиального отображения степени не больше $m$ (и соответственно ядра $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^m$).

Поэтому отображение ищем в виде
$\varphi : X \to Y, Y = \mathbb{R}^k, z \in Y, z_{(i)} = x_{(1)}^{i_1} x_{(2)}^{i_2} \dots x_{(n)}^{i_n}|_{i_1+i_2+\dots+i_n \le m}, 1 \le i \le k$ для
$X = \mathbb{R}^n, n = 2, m = 2$.
Ядро, соответствующее этому отображению: $K(x_1, x_2) = (\langle x_1, x_2 \rangle + 1)^2$.
Будем решать задачу

$$\begin{cases} -Q(\lambda) = \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j K(x_i, x_j) - \sum_{i=1}^m \lambda_i \to \min_\lambda, \\ 0 \le \lambda_i \le C, \\ \sum_{i=1}^m \lambda_i y_i = 0. \end{cases}$$

# Выбор подходящего ядра

## Указание [доказать]

Любые $m+1$ точек из $\mathbb{R}^n$ могут быть линейно разделены на любые два класса после перехода в новое пространство с помощью мономиального отображения степени не больше $m$ (и соответственно ядра $K(x_1, x_2) = (\langle x_1, x_2\rangle + 1)^m$).

Поэтому отображение ищем в виде
$\varphi : X \to Y, Y = \mathbb{R}^k, z \in Y, z_{(i)} = x_{(1)}^{i_1} x_{(2)}^{i_2} \dots x_{(n)}^{i_n}|_{i_1+i_2+\dots+i_n \le m}, 1 \le i \le k$ для
$X = \mathbb{R}^n, n = 2, m = 2.$
Ядро, соответствующее этому отображению: $K(x_1, x_2) = (\langle x_1, x_2\rangle + 1)^2$.
Будем решать задачу

$$\begin{cases} -Q(\lambda) = \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j K(x_i, x_j) - \sum_{i=1}^m \lambda_i \to \min_\lambda, \\ 0 \le \lambda_i \le C, \\ \sum_{i=1}^m \lambda_i y_i = 0. \end{cases}$$

**Упражнение.** Чему равна размерность $k$ для мономиального $\varphi : \mathbb{R}^n \to \mathbb{R}^k$?

# Необходимые расчеты 1

$$-Q(\lambda) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i\lambda_j y_i y_j(\langle x_i, x_j\rangle + 1)^2 - \sum_{i=1}^{3}\lambda_i =$$
$$\frac{1}{2}(2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 + \lambda_1^2 - 18\lambda_2\lambda_3 + 25\lambda_2^2 + 4\lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3).$$

# Необходимые расчеты 1

$$-Q(\lambda) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i\lambda_j y_i y_j(\langle x_i, x_j\rangle + 1)^2 - \sum_{i=1}^{3}\lambda_i =$$
$$\frac{1}{2}(2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 + \lambda_1^2 - 18\lambda_2\lambda_3 + 25\lambda_2^2 + 4\lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3).$$

Из условия $\sum_{i=1}^{3}\lambda_i y_i = 0$ имеем $\lambda_3 = \lambda_1 + \lambda_2$.

# Необходимые расчеты 1

$-Q(\lambda) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i\lambda_j y_i y_j(\langle x_i, x_j\rangle + 1)^2 - \sum_{i=1}^{3}\lambda_i =$
$\frac{1}{2}(2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 + \lambda_1^2 - 18\lambda_2\lambda_3 + 25\lambda_2^2 + 4\lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3).$
Из условия $\sum_{i=1}^{3}\lambda_i y_i = 0$ имеем $\lambda_3 = \lambda_1 + \lambda_2$.
Подставим в выражение для $-Q(\lambda) = \frac{1}{2}(3\lambda_1^2 + 11\lambda_2^2 - 10\lambda_1\lambda_2) - 2(\lambda_1 + \lambda_2)$.

# Необходимые расчеты 1

$-Q(\lambda) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i\lambda_j y_i y_j(\langle x_i, x_j\rangle + 1)^2 - \sum_{i=1}^{3}\lambda_i =$
$\frac{1}{2}(2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 + \lambda_1^2 - 18\lambda_2\lambda_3 + 25\lambda_2^2 + 4\lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3).$
Из условия $\sum_{i=1}^{3}\lambda_i y_i = 0$ имеем $\lambda_3 = \lambda_1 + \lambda_2$.
Подставим в выражение для $-Q(\lambda) = \frac{1}{2}(3\lambda_1^2 + 11\lambda_2^2 - 10\lambda_1\lambda_2) - 2(\lambda_1 + \lambda_2)$.
Дифференцируем $Q$ по $\lambda$:

$$\begin{cases} \frac{\partial Q}{\partial \lambda_1} = 0 & \Leftrightarrow 3\lambda_1 - 5\lambda_2 = 2, \\ \frac{\partial Q}{\partial \lambda_2} = 0 & \Leftrightarrow -5\lambda_1 + 11\lambda_2 = 2 \end{cases}$$

# Необходимые расчеты 1

$-Q(\lambda) = \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3}\lambda_i\lambda_j y_i y_j(\langle x_i, x_j\rangle + 1)^2 - \sum_{i=1}^{3}\lambda_i =$
$\frac{1}{2}(2\lambda_1\lambda_2 - 2\lambda_1\lambda_3 + \lambda_1^2 - 18\lambda_2\lambda_3 + 25\lambda_2^2 + 4\lambda_3^2) - (\lambda_1 + \lambda_2 + \lambda_3)$.
Из условия $\sum_{i=1}^{3}\lambda_i y_i = 0$ имеем $\lambda_3 = \lambda_1 + \lambda_2$.
Подставим в выражение для $-Q(\lambda) = \frac{1}{2}(3\lambda_1^2 + 11\lambda_2^2 - 10\lambda_1\lambda_2) - 2(\lambda_1 + \lambda_2)$.
Дифференцируем $Q$ по $\lambda$:

$$\begin{cases} \frac{\partial Q}{\partial \lambda_1} = 0 & \Leftrightarrow 3\lambda_1 - 5\lambda_2 = 2, \\ \frac{\partial Q}{\partial \lambda_2} = 0 & \Leftrightarrow -5\lambda_1 + 11\lambda_2 = 2 \end{cases}$$

Решая линейную систему уравнений, получаем $(\lambda_1, \lambda_2, \lambda_3) = (4, 2, 6)$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0,\ w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0,\ w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0,\ w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0$, $w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.
Теперь рассчитаем основную часть: $\sum_{i=1}^{3} \lambda_i y_i K(x_i, x) = 4 \cdot 1 + 2(2x_{(1)} + 1)^2 - 6(x_{(1)} + 1)^2$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0$, $w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.
Теперь рассчитаем основную часть: $\sum_{i=1}^{3} \lambda_i y_i K(x_i, x) = 4 \cdot 1 + 2(2x_{(1)} + 1)^2 - 6(x_{(1)} + 1)^2$.
Объединяя, получаем **разделяющую поверхность**: $f(x) = 2x_{(1)}^2 - 4x_{(1)} + 1$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0$, $w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.
Теперь рассчитаем основную часть: $\sum_{i=1}^{3} \lambda_i y_i K(x_i, x) = 4 \cdot 1 + 2(2x_{(1)} + 1)^2 - 6(x_{(1)} + 1)^2$.
Объединяя, получаем **разделяющую поверхность**: $f(x) = 2x_{(1)}^2 - 4x_{(1)} + 1$.

**Нули разделителя**: $f(x) = 0 \Leftrightarrow x_{(1)} = 1 \pm \frac{\sqrt{2}}{2}$.

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0$, $w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.
Теперь рассчитаем основную часть: $\sum_{i=1}^{3} \lambda_i y_i K(x_i, x) = 4 \cdot 1 + 2(2x_{(1)} + 1)^2 - 6(x_{(1)} + 1)^2$.
Объединяя, получаем **разделяющую поверхность**: $f(x) = 2x_{(1)}^2 - 4x_{(1)} + 1$.

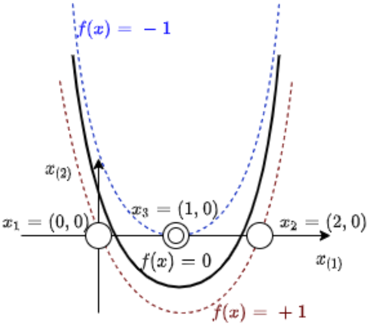
**Нули разделителя**: $f(x) = 0 \Leftrightarrow x_{(1)} = 1 \pm \frac{\sqrt{2}}{2}$.
**Первый край полосы** $(f(x) = -1)$:
$f_1(x) = f(x) + 1 = 2(x_{(1)} - 1)^2$.
**Нули**: $x_{(1)} = 1$ (это — исходная точка $x_3$).

# Необходимые расчеты 2

Найдем разделяющую поверхность в виде:
$f(x) = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x) - w_0,\ w_0 = \sum_{i=1}^{3} \lambda_i y_i K(x_i, x_j) - y_j, x = (x_{(1)}, x_{(2)})$.
Поскольку все $\lambda > 0$, то можем взять в качестве опорного вектора $x_1$.
Получим $w_0 = 4K(x_1, x_1) + 2K(x_2, x_1) - 6K(x_3, x_1) - 1 = -1$.
Теперь рассчитаем основную часть: $\sum_{i=1}^{3} \lambda_i y_i K(x_i, x) = 4 \cdot 1 + 2(2x_{(1)} + 1)^2 - 6(x_{(1)} + 1)^2$.
Объединяя, получаем **разделяющую поверхность**: $f(x) = 2x_{(1)}^2 - 4x_{(1)} + 1$.

**Нули разделителя**: $f(x) = 0 \Leftrightarrow x_{(1)} = 1 \pm \frac{\sqrt{2}}{2}$.
**Первый край полосы** $(f(x) = -1)$:
$f_1(x) = f(x) + 1 = 2(x_{(1)} - 1)^2$.
**Нули**: $x_{(1)} = 1$ (это — исходная точка $x_3$).
**Второй край полосы** $(f(x) = +1)$:
$f_2(x) = f(x) - 1 = 2x_{(1)}(x_{(1)} - 2)$.
**Нули**: $x_{(1)} = 0, x_{(1)} = 2$ (это — исходные точки $x_1, x_2$).

Построим визуализацию на плоскости как разделяющей поверхности, так и двух краев полосы:



# Время для вопросов

# Источники

На основе материалов сайта http://www.machinelearning.ru.