

# Машинное обучение

Лекция 7. Линейные модели, SVM

МФТИ 2018

Материалы: В. Кантор, К. Воронцов

# План

1. Логистическая регрессия
2. Метод Опорных Векторов (SVM)
3. О применении линейных моделей

# 1. Логистическая регрессия

# Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

# Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^N L(y_i, a(x_i))$$

## Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^N L(y_i, a(x_i))$$

$$L(y_i, a(x_i)) = (y_i - a(x_i))^2$$

## Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^N L(y_i, a(x_i))$$

$$L(y_i, a(x_i)) = (y_i - a(x_i))^2$$

$$L(y_i, a(x_i)) = |y_i - a(x_i)|$$

## Линейная классификация

$$a(x) = \begin{cases} 1, & \text{если } f(x) > 0 \\ -1, & \text{если } f(x) \leq 0 \end{cases}$$

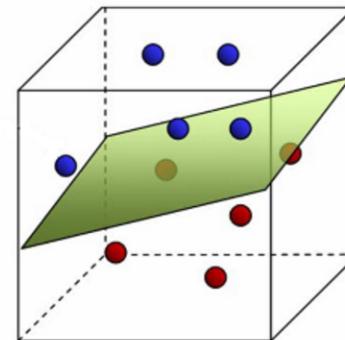
$$f(x) = w_0 + w_1x_1 + \cdots + w_nx_n$$

# Линейная классификация

$$a(x) = \begin{cases} 1, & \text{если } f(x) > 0 \\ -1, & \text{если } f(x) \leq 0 \end{cases}$$

$$f(x) = w_0 + w_1x_1 + \cdots + w_nx_n = w_0 + \langle w, x \rangle$$

Геометрическая интерпретация:  
разделяем классы плоскостью

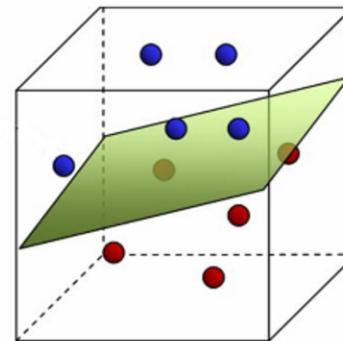


# Линейная классификация

$$a(x) = \begin{cases} 1, & \text{если } f(x) > 0 \\ -1, & \text{если } f(x) \leq 0 \end{cases}$$

$$f(x) = \langle w, x \rangle$$

Геометрическая интерпретация:  
разделяем классы плоскостью



## Отступ (margin)

Отступом алгоритма  $a(x) = \text{sign}\{f(x)\}$  на объекте  $x_i$  называется величина  $M_i = y_i f(x_i)$   
( $y_i$  - класс, к которому относится  $x_i$ )

$$\begin{aligned}M_i \leq 0 &\Leftrightarrow y_i \neq a(x_i) \\M_i > 0 &\Leftrightarrow y_i = a(x_i)\end{aligned}$$

## Функция потерь

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0]$$

## Функция потерь

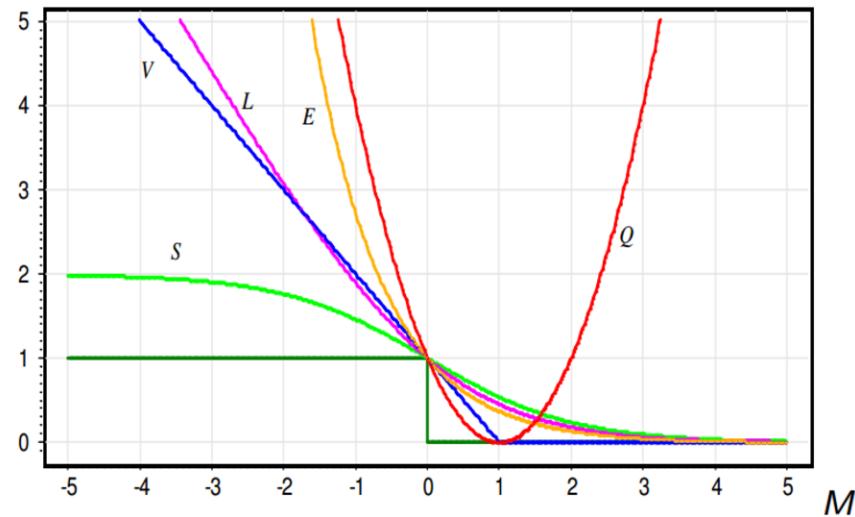
$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leq \tilde{Q}(w) = \sum_{i=1}^{\ell} \mathcal{L}(M_i(w)) \rightarrow \min_w;$$

Функция эмпирического риска

Функция потерь

# ФУНКЦИЯ ПОТЕРЬ

$$Q(w) = \sum_{i=1}^{\ell} [M_i(w) < 0] \leq \tilde{Q}(w) = \sum_{i=1}^{\ell} \mathcal{L}(M_i(w)) \rightarrow \min_w;$$



$$\begin{aligned} Q(M) &= (1 - M)^2 \\ V(M) &= (1 - M)_+ \\ S(M) &= 2(1 + e^M)^{-1} \\ L(M) &= \log_2(1 + e^{-M}) \\ E(M) &= e^{-M} \end{aligned}$$

## Логистическая регрессия

$$y_i \in \{0, 1\} \quad Q = - \sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \rightarrow \min_w$$
$$p_i = \sigma(-\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}}$$

## Логистическая регрессия

$$y_i \in \{0, 1\} \quad Q = - \sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \rightarrow \min_w$$

$$p_i = \sigma(\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}} = P(y = 1|x)$$

## Логистическая регрессия

$$y_i \in \{0, 1\} \quad Q = - \sum_{i=1}^{\ell} y_i \ln p_i + (1 - y_i) \ln(1 - p_i) \rightarrow \min_w$$
$$p_i = \sigma(\langle w, x_i \rangle) = \frac{1}{1 + e^{-\langle w, x_i \rangle}} = P(y = 1 | x)$$

Как правило, добавляется  $\ell_1$  или  $\ell_2$ -регуляризация, а оптимизационная задача решается с помощью SGD или метода Ньютона-Рафсона

## Эквивалентность оптимизационных задач

$$Q = - \sum_{i=1}^{\ell} y_i \ln \frac{1}{1 + e^{-\langle w, x_i \rangle}} + (1 - y_i) \ln \frac{1}{1 + e^{\langle w, x_i \rangle}} \rightarrow \min_w$$

$$-y_i \ln \frac{1}{1 + e^{-\langle w, x_i \rangle}} - (1 - y_i) \ln \frac{1}{1 + e^{\langle w, x_i \rangle}} = \begin{cases} \ln(1 + e^{-\langle w, x_i \rangle}), & y_i = 1 \\ \ln(1 + e^{\langle w, x_i \rangle}), & y_i = 0 \end{cases}$$

$$Q = \sum_{i=1}^{\ell} \underbrace{\ln(1 + e^{-y_i \langle w, x_i \rangle})}_{L(M) = \ln(1 + e^{-M_i})} \rightarrow \min_w \quad y_i \in \{-1, 1\}$$

## 2. Метод Опорных Векторов (SVM)

# Метод опорных векторов

## Линейный классификатор:

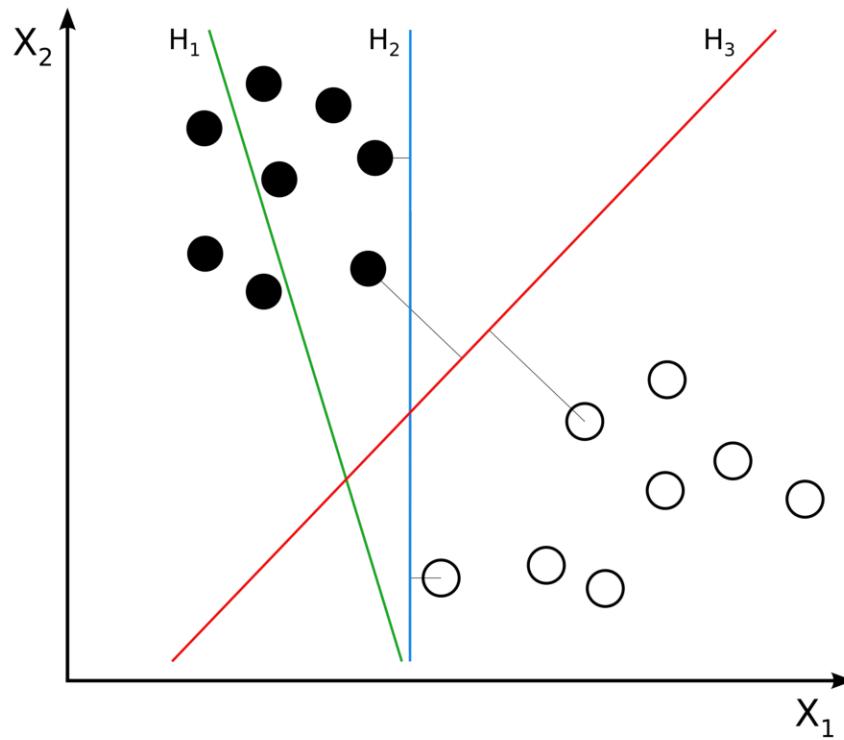
$$a(x) = sign(< w, x > - w_0)$$

Использующий кусочно-линейную функцию потерь и L2-регуляризатор:

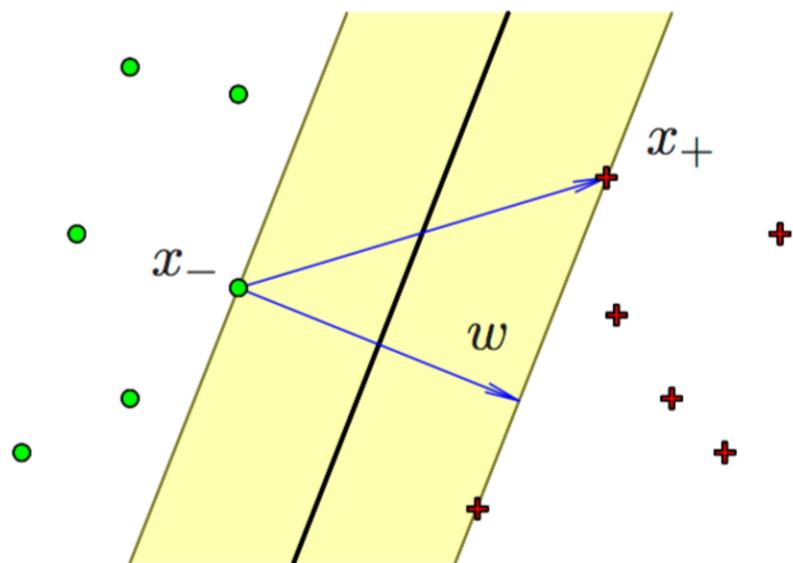
кусочно-линейная функция потерь:

$$L(M_i) = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

# Построение разделяющей гиперплоскости

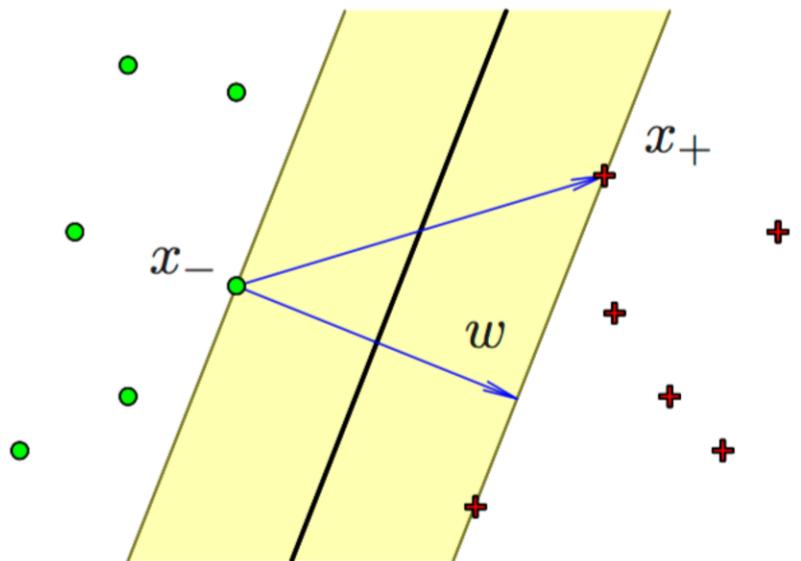


## Разделяющая полоса



$$a(x) = \operatorname{sign}\left(\sum_{j=1}^n w_j x^j - w_0\right) = \operatorname{sign}(\langle w, x \rangle - w_0)$$

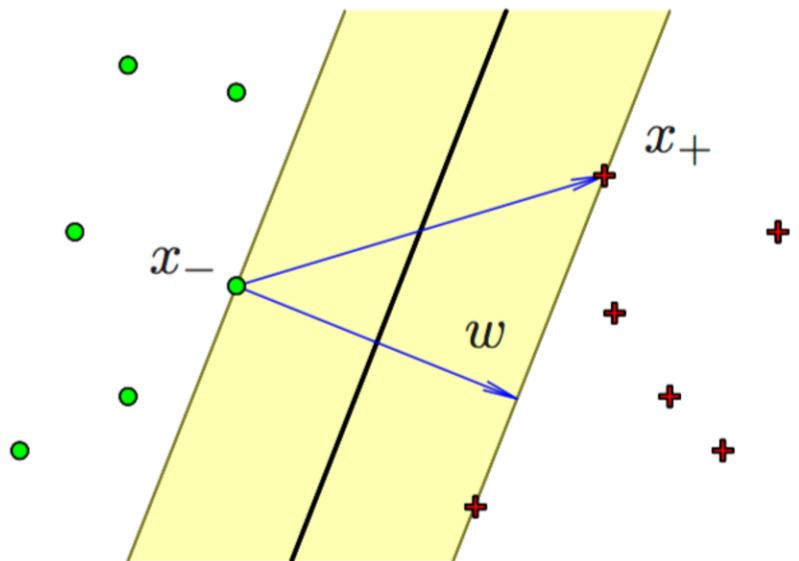
## Разделяющая полоса



$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1$$

$$a(x) = \operatorname{sign}\left(\sum_{j=1}^n w_j x^j - w_0\right) = \operatorname{sign}(\langle w, x \rangle - w_0)$$

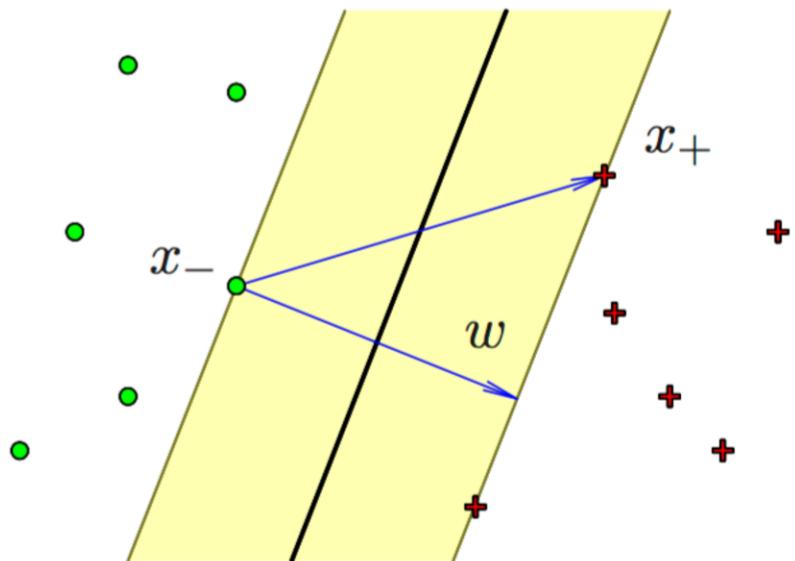
## Ширина разделяющей полосы



$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle$$

$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1$$

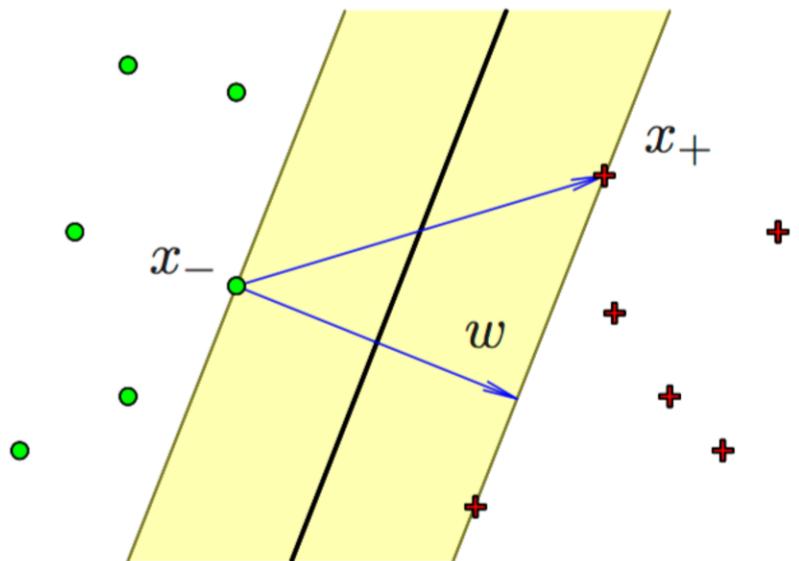
## Ширина разделяющей полосы



$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+ \rangle - \langle w, x_- \rangle}{\|w\|}$$

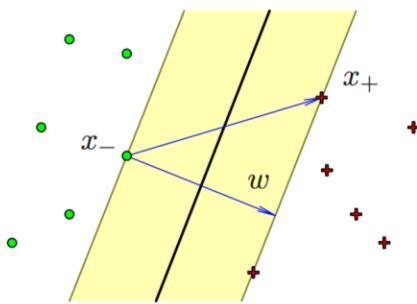
# Ширина разделяющей полосы



$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+ \rangle - \langle w, x_- \rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

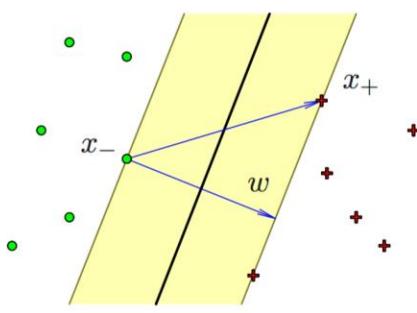
# Ширина разделяющей полосы



$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1.$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+ \rangle - \langle w, x_- \rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

# Максимизация зазора

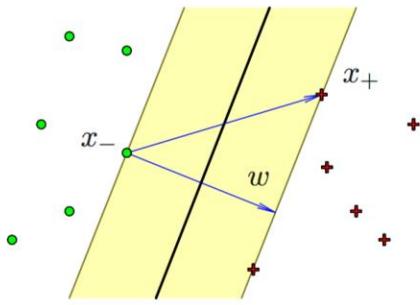


$$\min_{i=1, \dots, \ell} y_i (\langle w, x_i \rangle - w_0) = 1$$

$$\left\langle (x_+ - x_-), \frac{w}{\|w\|} \right\rangle = \frac{\langle w, x_+ \rangle - \langle w, x_- \rangle}{\|w\|} = \frac{(w_0 + 1) - (w_0 - 1)}{\|w\|} = \frac{2}{\|w\|}$$

$$\begin{cases} \langle w, w \rangle \rightarrow \min; \\ y_i (\langle w, x_i \rangle - w_0) \geq 1, \quad i = 1, \dots, \ell. \end{cases}$$

## Случай линейно неразделимой выборки



$$\begin{cases} \langle w, w \rangle \rightarrow \min; \\ y_i(\langle w, x_i \rangle - w_0) \geq 1, \quad i = 1, \dots, \ell. \end{cases}$$

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

## Оптимизационная задача в SVM

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

Причем здесь линейный классификатор  
в привычном нам виде?

## Безусловная оптимизационная задача в SVM

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

Напоминание:

$$M_i = y_i (\langle w, x_i \rangle - w_0)$$

отступ на  $i$ -том объекте

## Безусловная оптимизационная задача в SVM

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

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i$$

$$\sum_{i=1}^l \xi_i \rightarrow \min$$

Напоминание:

$$M_i = y_i (\langle w, x_i \rangle - w_0)$$

отступ на  $i$ -том объекте

## Безусловная оптимизационная задача в SVM

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

Напоминание:

$M_i = y_i (\langle w, x_i \rangle - w_0)$   
отступ на  $i$ -том объекте

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i \quad \Rightarrow \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^l \xi_i \rightarrow \min$$

## Безусловная оптимизационная задача в SVM

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

Напоминание:

$$M_i = y_i (\langle w, x_i \rangle - w_0)$$

отступ на  $i$ -том объекте

$$\xi_i \geq 0$$

$$\xi_i \geq 1 - M_i \quad \Rightarrow \xi_i = \max\{0, 1 - M_i\} = (1 - M_i)_+$$

$$\sum_{i=1}^l \xi_i \rightarrow \min$$

$$Q(w, w_0) = \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+ + \frac{1}{2C} \|w\|^2 \rightarrow \min_{w, w_0}$$

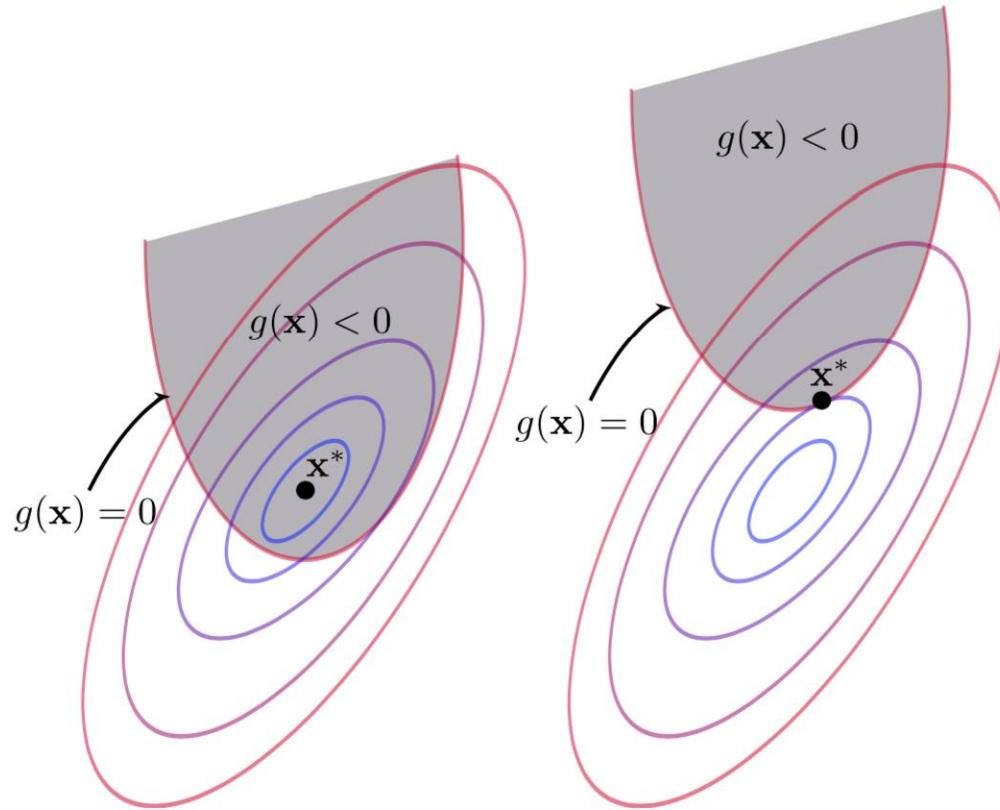
## Безусловная оптимизационная задача в SVM

$$Q(w, w_0) = \sum_{i=1}^{\ell} (1 - M_i(w, w_0))_+ + \frac{1}{2C} \|w\|^2 \rightarrow \min_{w, w_0}$$

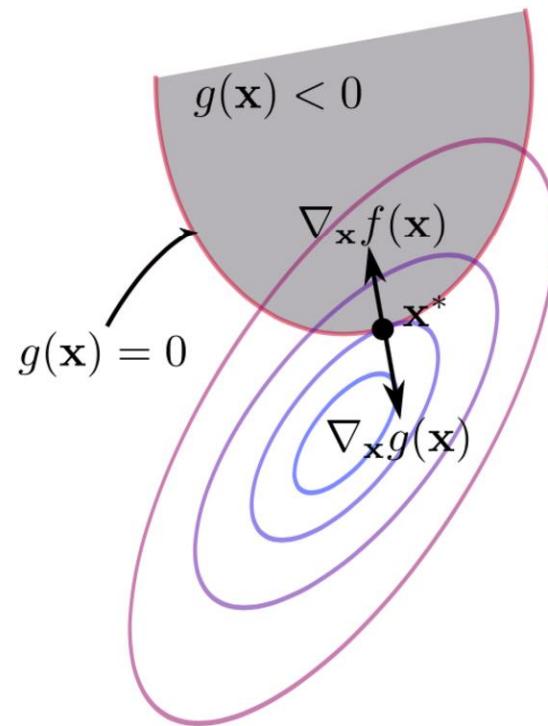
Напоминание: теорема Каруша-Куна-Таккера

$$\begin{cases} f(x) \rightarrow \min_x \\ g(x) \leq 0 \end{cases} \implies \begin{cases} \nabla_x L(x, \mu) = \nabla_x(f(x) + \mu g(x)) = 0 \\ \mu g(x) = 0 \\ \mu \geq 0 \end{cases}$$

## Напоминание: теорема Каруша-Куна-Таккера



Напоминание: теорема Каруша-Куна-Таккера



## Двойственная задача в SVM

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

## Двойственная задача в SVM

$$\begin{cases} \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, w_0, \xi}; \\ y_i (\langle w, x_i \rangle - w_0) \geq 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{cases} \quad M_i(w, w_0) = y_i (\langle w, x_i \rangle - w_0)$$

## Двойственная задача в SVM

$$\begin{cases} \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{w, w_0, \xi}; \\ y_i (\langle w, x_i \rangle - w_0) \geq 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{cases} \quad M_i(w, w_0) = y_i (\langle w, x_i \rangle - w_0)$$

$$\begin{aligned} \mathcal{L}(w, w_0, \xi; \lambda, \eta) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1 + \xi_i) - \sum_{i=1}^{\ell} \xi_i \eta_i = \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1) - \sum_{i=1}^{\ell} \xi_i (\lambda_i + \eta_i - C), \end{aligned}$$

## Двойственная задача в SVM

$$\begin{aligned}\mathcal{L}(w, w_0, \xi; \lambda, \eta) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1 + \xi_i) - \sum_{i=1}^{\ell} \xi_i \eta_i = \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1) - \sum_{i=1}^{\ell} \xi_i (\lambda_i + \eta_i - C),\end{aligned}$$

$$\left\{ \begin{array}{l} \mathcal{L}(w, w_0, \xi; \lambda, \eta) \rightarrow \min_{w, w_0, \xi} \max_{\lambda, \eta}; \\ \xi_i \geq 0, \quad \lambda_i \geq 0, \quad \eta_i \geq 0, \quad i = 1, \dots, \ell; \\ \lambda_i = 0 \text{ либо } M_i(w, w_0) = 1 - \xi_i, \quad i = 1, \dots, \ell; \\ \eta_i = 0 \text{ либо } \xi_i = 0, \quad i = 1, \dots, \ell; \end{array} \right.$$

# Опорные векторы в SVM

$$\begin{aligned}\mathcal{L}(w, w_0, \xi; \lambda, \eta) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{\ell} \xi_i - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1 + \xi_i) - \sum_{i=1}^{\ell} \xi_i \eta_i = \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^{\ell} \lambda_i (M_i(w, w_0) - 1) - \sum_{i=1}^{\ell} \xi_i (\lambda_i + \eta_i - C),\end{aligned}$$

$$\frac{\partial \mathcal{L}}{\partial w} = w - \sum_{i=1}^{\ell} \lambda_i y_i x_i = 0 \quad \implies \quad w = \sum_{i=1}^{\ell} \lambda_i y_i x_i;$$

## Двойственная задача: итог

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = -\lambda_i - \eta_i + C = 0 \quad \Rightarrow \quad \eta_i + \lambda_i = C, \quad i = 1, \dots, \ell.$$

$$\begin{cases} -\mathcal{L}(\lambda) = -\sum_{i=1}^{\ell} \lambda_i + \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle \rightarrow \min_{\lambda}; \\ 0 \leq \lambda_i \leq C, \quad i = 1, \dots, \ell; \\ \sum_{i=1}^{\ell} \lambda_i y_i = 0. \end{cases}$$

# Двойственная задача: итог

$$a(x) = \operatorname{sign} \left( \sum_i \lambda_i y_i \langle x_i, x \rangle - w_0 \right)$$

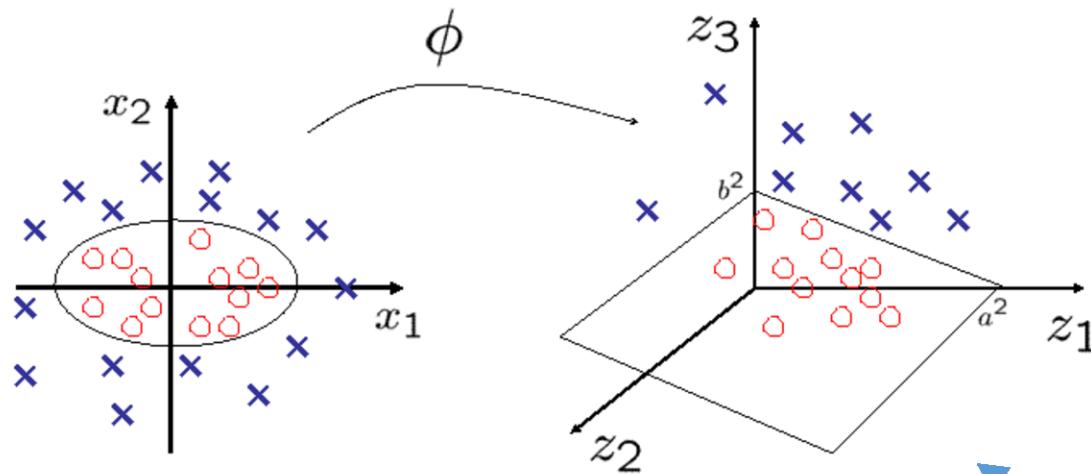
$$w = \sum_i \lambda_i y_i x_i$$

$$w_0 = \operatorname{Med}(y_i - \langle w, x_i \rangle \mid \lambda_i > 0)$$

## Ключевые моменты

- Метод опорных векторов – линейный классификатор с кусочно-линейной функцией потерь (hinge loss) и L2-регуляризатором
- Придуман метод был из соображений максимизации зазора между классами
- В случае линейно разделимой выборки это означает просто максимизацию ширины разделяющей полосы
- А в случае линейно неразделимой выборки просто добавляется возможность попадания объектов в полосу и штрафы за эти попадания

# Добавление новых признаков



$$\phi : (x_1, x_2) \longrightarrow (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$\left(\frac{x_1}{a}\right)^2 + \left(\frac{x_2}{b}\right)^2 = 1 \longrightarrow \frac{z_1}{a^2} + \frac{z_3}{b^2} = 1$$

Спрямляющее  
пространство

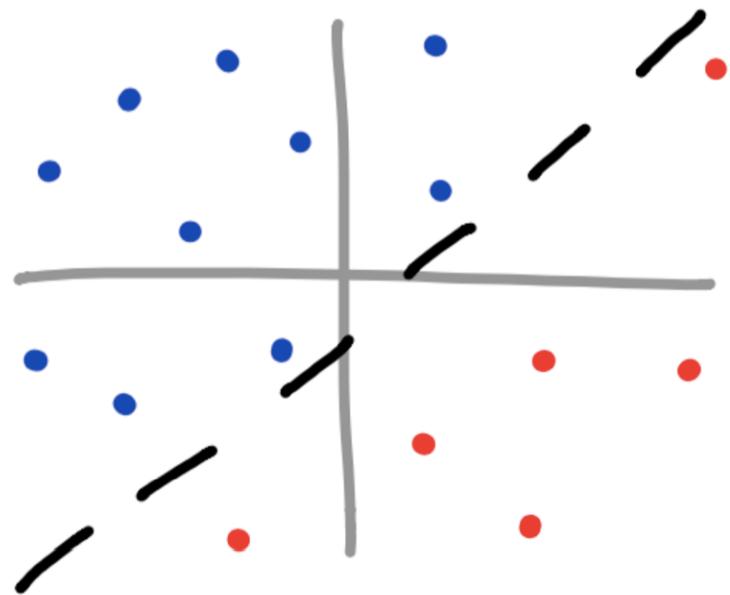
## Kernel Trick

$$\begin{aligned}x &\mapsto \phi(x) \\w &\mapsto \phi(w)\end{aligned}\Rightarrow \langle w, x \rangle \mapsto \langle \phi(w), \phi(x) \rangle$$

Можно не делать преобразование признаков явно, а вместо скалярного произведения  $\langle w, x \rangle$  использовать функцию  $K(w, x)$ , представимую в виде:

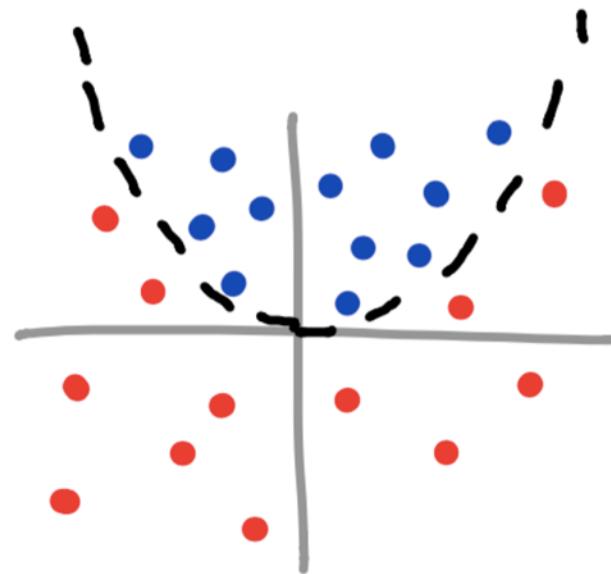
$$K(w, x) = \langle \phi(w), \phi(x) \rangle$$

## Линейное ядро



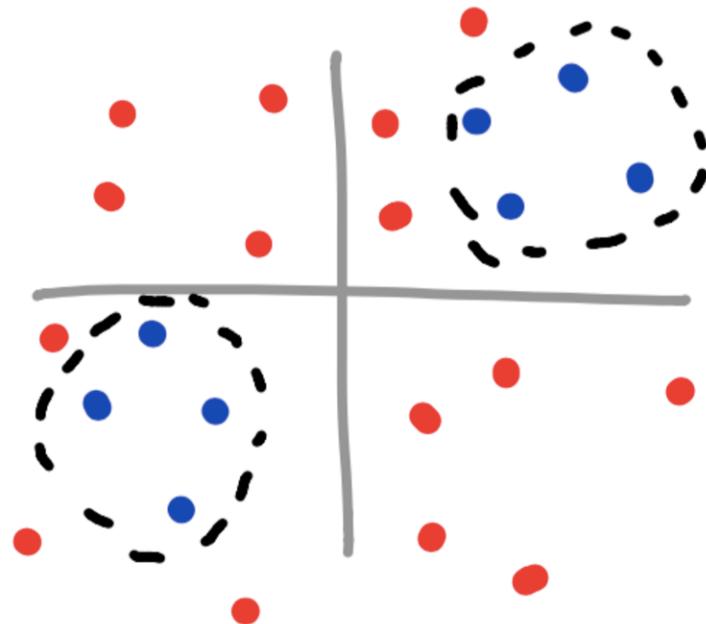
$$K(w, x) = \langle w, x \rangle$$

## Полиномиальное ядро



$$K(w, x) = (\gamma \langle w, x \rangle + r)^d$$

## Радиальное ядро



$$K(w, x) = e^{-\gamma \|w-x\|^2}$$

# Ядра и библиотеки

- LibSVM – можно выбирать ядра
- LibLinear – только линейное ядро
- Scikit-learn – обертка над LibSVM и LibLinear
- Vowpal Wabbit – только линейное ядро

## Вид классификатора в SVM

$$a(x) = \text{sign}\left(\sum_{i=1}^{\ell} \lambda_i y_i \langle x_i, x \rangle - w_0\right)$$

Вид классификатора в SVM: теперь с ядром

$$a(x) = \text{sign} \left( \sum_{i=1}^h \lambda_i y_i K(x_i, x) - w_0 \right)$$

# Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим  $K$  бинарных классификаторов
- One-to-One
  - Строим  $K * (K - 1)/2$  классификаторов
- Multiclass-SVM

# Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

# Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим  $K$  бинарных классификаторов

# Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим  $K$  бинарных классификаторов
- One-to-One
  - Строим  $K * (K - 1)/2$  классификаторов

# Мультиклассовая задача классификации

Как адаптировать SVM под задачу мультиклассовой классификации?

$$Y = \{1, \dots, K\}$$

- One-to-All
  - Строим  $K$  бинарных классификаторов
- One-to-One
  - Строим  $K * (K - 1)/2$  классификаторов
- Multiclass-SVM

# Мультиклассовая задача: One-All Approach

In case of multilabel classification

$$y \in \{1, \dots, K\}$$

Lets train  $K$  classifiers

$$a_k(x) = \text{sign}(\langle w_k, x \rangle + b_k),$$

$k$ -th training sample is  $(x_i, 2[y_i = k] - 1)_{i=1}^l$ ,  $k = 1, 2, \dots, K$ .

Final classifier:

$$a(x) = \arg \max_{k \in \{1, \dots, K\}} \left\{ \langle w_k, x \rangle + b_k \right\}.$$

# Мультиклассовая задача: One-to-One

Lets train  $C_K^2$  classifiers

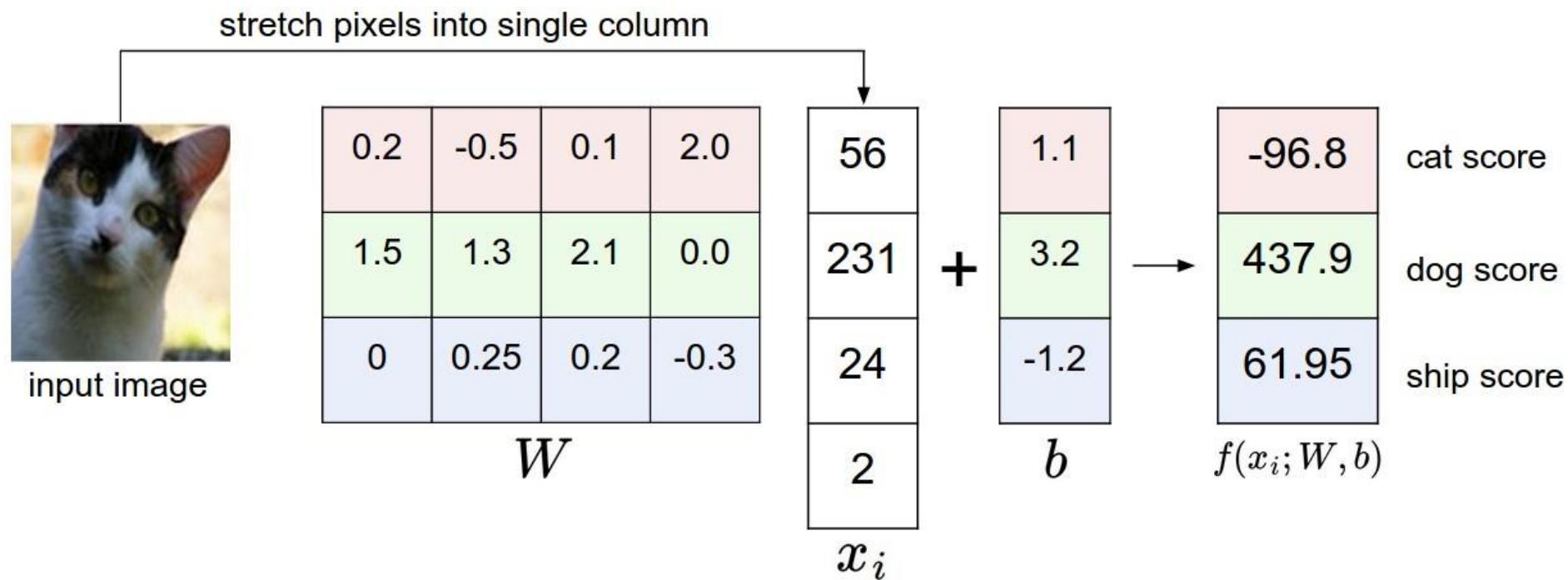
$$a_{ij}(x), \quad k = 1, 2, \dots, K, \quad i \neq j$$

Training sample is set of objects with labels  $i$  and  $j$ .

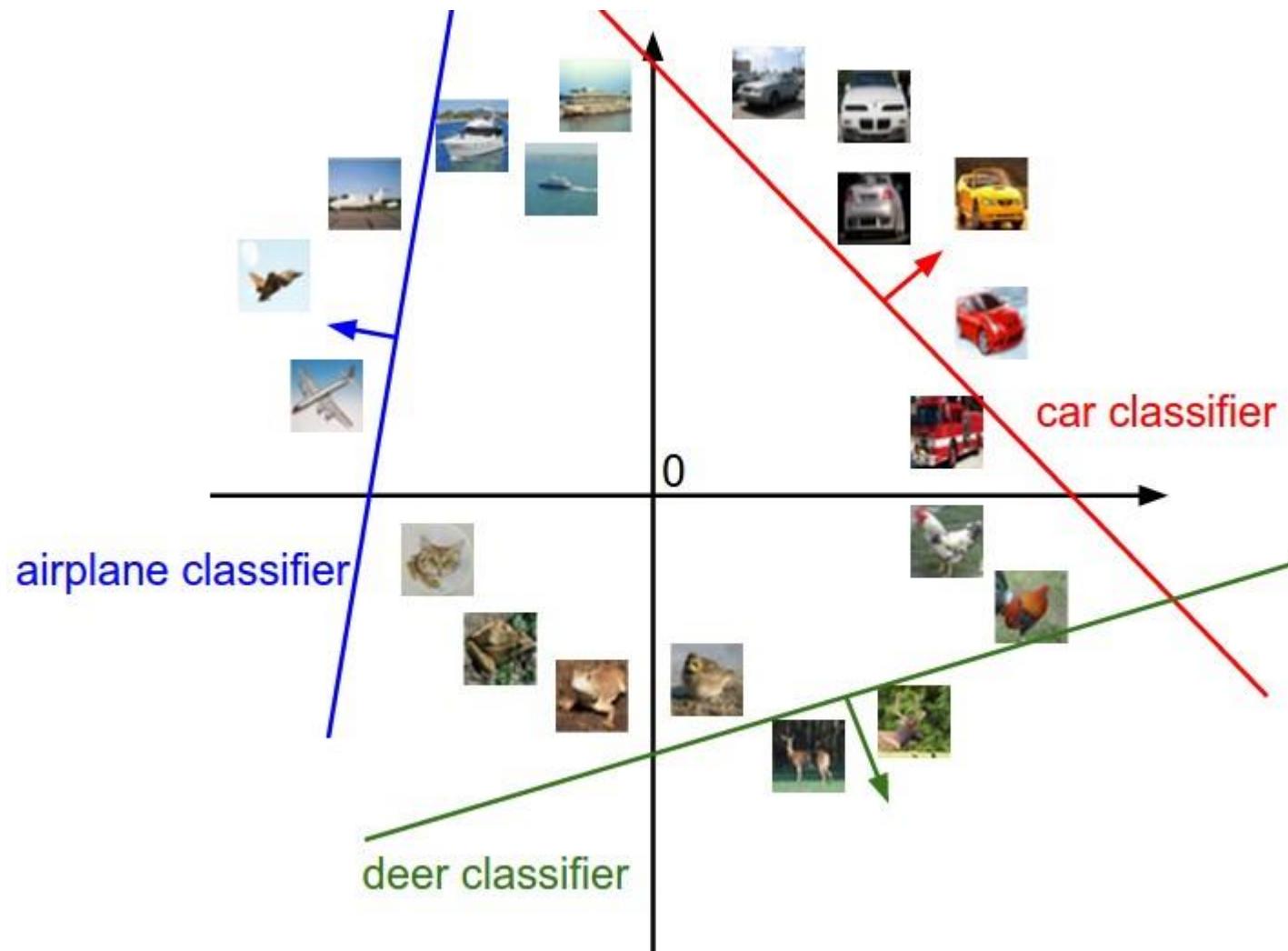
Final classifier:

$$a(x) = \arg \max_{k \in \{1, \dots, K\}} \sum_{i=1}^{\ell} \sum_{j \neq i} [a_{ij}(x) = k].$$

# Мультиклассовая задача: Multiclass-SVM



# Мультиклассовая задача: Multiclass-SVM



# Мультиклассовая задача: Multiclass-SVM

Lets fit  $K$  sets of parameters  $w_1, \dots, w_K$ .

Final classifier is defined as:

$$a(x) = \arg \max_{k \in \{1, \dots, K\}} \langle w_k, x \rangle.$$

Lets define loss function as:

$$\max_k \left\{ \langle w_k, x \rangle + 1 - [k = y(x)] \right\} - \langle w_{y(x)}, x \rangle.$$

Margin in multilabel case is Frobenius norm of  $W \in \mathbb{R}^{K \times n}$ :

$$\rho = \frac{1}{\|W\|^2} = \frac{1}{\sum_{k=1}^K \sum_{j=1}^d w_{kj}^2}.$$

# Мультиклассовая задача: Multiclass-SVM

In linear separable case:

$$\begin{cases} \frac{1}{2} \|W\|^2 \rightarrow \min_W \\ \langle w_{y_i}, x_i \rangle + [y_i = k] - \langle w_k, x_i \rangle \geq 1, \quad i = 1, \dots, \ell; k = 1, \dots, K. \end{cases}$$

In inseparable case:

$$\begin{cases} \frac{1}{2} \|W\|^2 + C \sum_{i=1}^{\ell} \xi_i \rightarrow \min_{W, \xi} \\ \langle w_{y_i}, x_i \rangle + [y_i = k] - \langle w_k, x_i \rangle \geq 1 - \xi_i, \quad i = 1, \dots, \ell; k = 1, \dots, K; \\ \xi_i \geq 0, \quad i = 1, \dots, \ell. \end{cases}$$

# SVM для регрессии

# Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

# Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^N L(y_i, a(x_i))$$

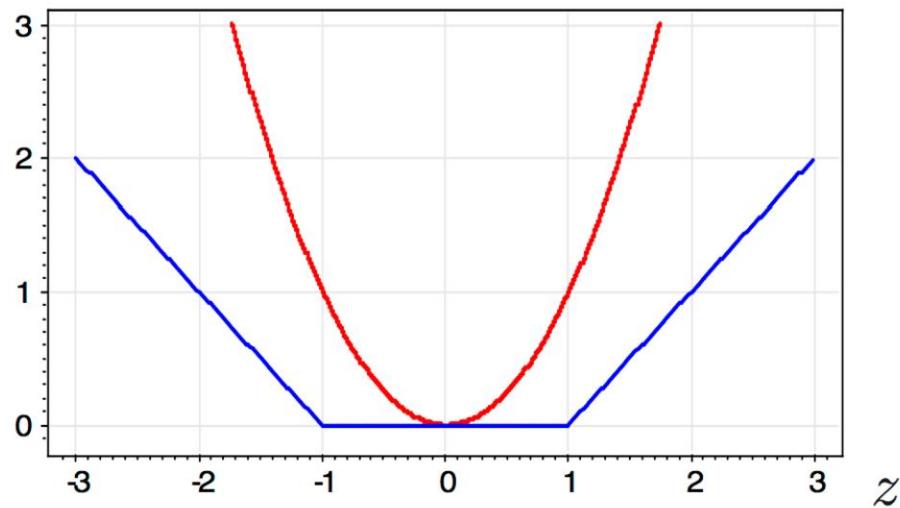
# Линейная регрессия

$$a(x) = \langle w, x \rangle + w_0$$

$$Q = \sum_{i=1}^N L(y_i, a(x_i))$$

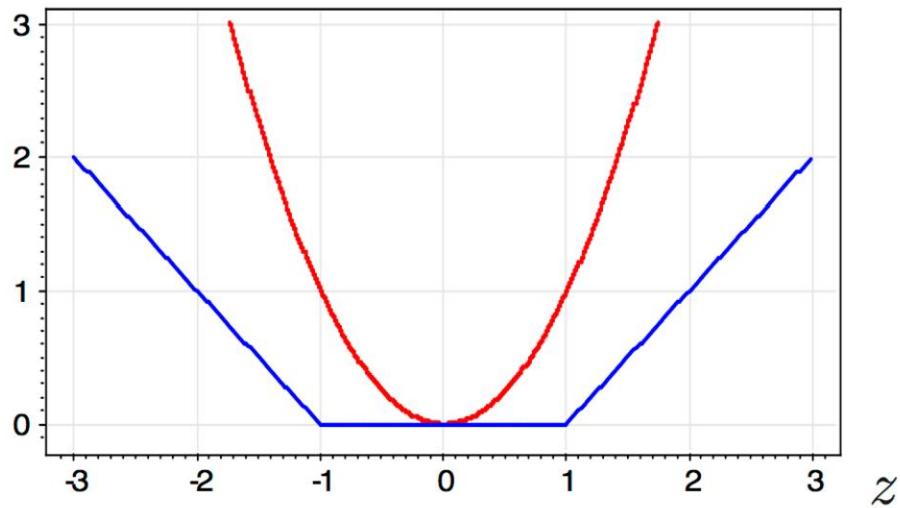
$$L(y_i, a(x_i)) = (y_i - a(x_i))^2$$

# Линейная регрессия



$$Q(a, X^\ell) = \sum_{i=1}^{\ell} (\langle w, x_i \rangle - w_0 - y_i)^2 + \tau \|w\|^2 \rightarrow \min_{w, w_0}$$

# SVM для регрессии



$$Q_\varepsilon(a, X^\ell) = \sum_{i=1}^{\ell} |\langle w, x_i \rangle - w_0 - y_i|_\varepsilon + \tau \langle w, w \rangle^2 \rightarrow \min_{w, w_0}$$

### 3. О применении линейных моделей

# Классификация текстов

The screenshot shows the 'All About the Company' section of the TOTAL website. At the top left is a logo with the text 'the world of TOTAL'. To the right is a navigation menu with links like 'All About The Company', 'Global Activities', 'Corporate Structure', 'TOTAL's Story', 'Upstream Strategy', 'Downstream Strategy', 'Chemicals Strategy', 'TOTAL Foundation', and 'Homepage'. Below the menu is a large image of a modern office building with the word 'TOTAL' on it, and a small bird flying above it. The main heading 'all about the company' is centered below the image. A paragraph of text follows, mentioning global operations in over 100 countries. Another paragraph discusses the company's strength from fast-growing oil and gas reserves, emphasizing natural gas. A third paragraph talks about expanding refining and marketing operations in Asia and the Mediterranean Basin. A final paragraph mentions the growing specialty chemicals sector. At the bottom of the page is a horizontal dotted line.



aardvark	0
about	2
all	2
Africa	1
apple	0
anxious	0
...	
gas	1
...	
oil	1
...	
Zaire	0

# Обработка разреженных признаков

Часть признаков – разреженные, а часть – плотные,  
как быть?

- Строим на разреженных признаках линейную модель
- Ответ линейной модели добавляем как еще один плотный признак
- Либо строим на плотных признаках другую модель и усредняем с ней

# Простые модели сложных систем

- Прогнозирование охвата целевой аудитории по параметрам рекламной кампании
- Прогнозирование дефектов в продукции на заводе
- Прогнозирование хим. состава в результате сложного процесса производства

*Как правило – везде сложные зависимости от большого числа параметров, но не все параметры известны и мало точек*

# Резюме

1. Логистическая регрессия имеет яркую вероятностную интерпретацию
2. SVM – лучший метод линейной классификации
3. SVM обобщается для нелинейной классификации для линейной и нелинейной регрессии
4. Легко обобщается для задачи регрессии
5. Неясно, как подбираться ядро для реальных задач

Отзывы о лекции: <https://goo.gl/forms/zeZiu1fSgrpPGp6T2>