

# Метод опорных векторов (Support Vector Machine — SVM)

Неделько В. М.

Институт математики СО РАН, г. Новосибирск  
nedelko@math.nsc.ru

Спецкурс «Теория статистических решений».  
Лекция 4.

# Общая характеристика

Метод обобщённого портрета (60–70-е годы, В.Н. Вапник и др.): отдалить объекты от разделяющей поверхности.  
В 90-е годы метод стал называться машиной опорных векторов (support vector machine, SVM).

Свойства:

- сводится к эффективно решаемой задаче квадратичного программирования,
- разреженность (решение определяется опорными векторами),
- обобщается введением функции ядра.

# Основные понятия

$X = R^n$  – пространство значений прогнозирующих переменных,

$Y = \{-1, 1\}$  – прогнозируемая переменная,

$D = X \times Y$ .

Решающая функция (алгоритм классификации)

$$f : X \rightarrow Y.$$

$V = \{(x^i, y^i) \in D \mid i = \overline{1, N}\}$  – случайная независимая выборка,  $V \in D^N$ .

$Q: D^N \rightarrow \Phi$  — метод построения решающих функций,  
 $\Phi$  — заданный класс решающих функций.

# Постановка задачи

Эмпирический риск:

$$\tilde{R}(V, f) = \frac{1}{N} \sum_{i=1}^N I(y^i \neq f(x^i)),$$

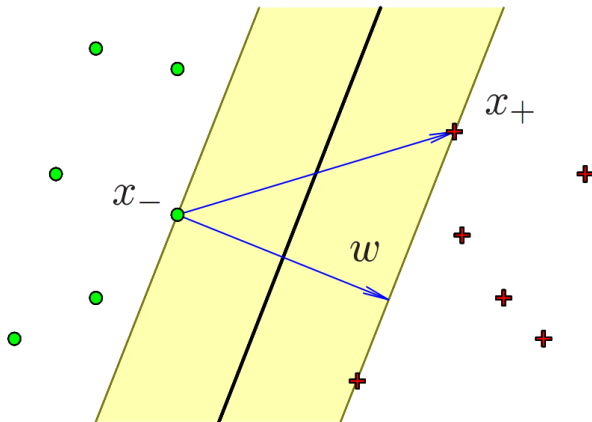
где  $I(\cdot)$  – индикаторная функция.

Линейный пороговый классификатор

$$f(x) = \text{sign}(wx - w_0).$$

Требуется найти вектор  $w$  и скаляр  $w_0$ , минимизирующие эмпирический риск при дополнительных требованиях.

# Линейно разделимая выборка



# Критерий

Нормировка

$$\min_{(x^i, y^i) \in V} y^i (x^i w - w_0) = 1.$$

Для граничных точек

$$x_+ w - w_0 = 1, \quad -(x_- w - w_0) = 1.$$

Ширина разделяющей полосы

$$(x_+ - x_-) \cdot \frac{w}{|w|} = \frac{(w_0 + 1) - (w_0 - 1)}{|w|} = \frac{2}{|w|}.$$

# Оптимизационная задача

Задача квадратичной оптимизации

$$\begin{cases} w^2 \rightarrow \min_{w, w_0} \\ y^i(x^i w - w_0) \geq 1, \quad i = 1, \dots, N. \end{cases}$$

Условие нормировки выполняется автоматически.

# Линейно неразделимая выборка

Задача квадратичного программирования

$$\begin{cases} \frac{w^2}{2} + C \sum_{i=1}^N \xi_i \rightarrow \min_{w, w_0, \xi} \\ y^i(x^i w - w_0) \geq 1 - \xi_i, \quad i = 1, \dots, N, \\ \xi_i \geq 0, \quad i = 1, \dots, N, \end{cases}$$

где  $C > 0$  – параметр.



# Функция Лагранжа

Задача на нахождение экстремума

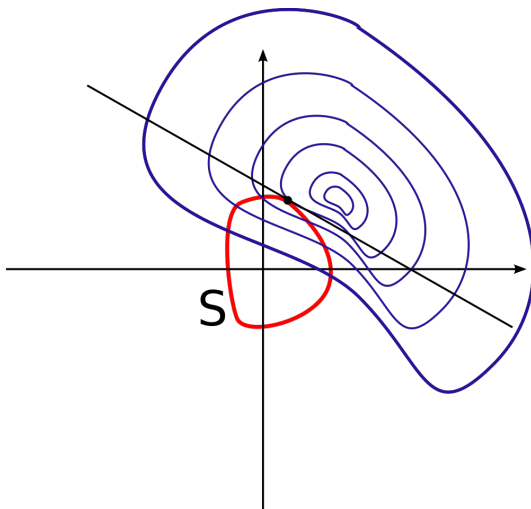
$$\begin{cases} f(x) \rightarrow \min_x \\ \psi(x) = 0. \end{cases}$$

Условие касания линий уровня

$$\frac{\partial f(x)}{\partial x} = \lambda \frac{\partial \psi(x)}{\partial x}.$$

После переноса в левую часть получаем функцию Лагранжа.

## Линии уровня и ограничение



# Условия Каруша–Куна–Таккера

Задача на нахождение экстремума

$$\begin{cases} f(x) \rightarrow \min_x \\ \psi(x) \geq 0. \end{cases}$$

Необходимые условия экстремума

$$\begin{cases} \frac{\partial f(x)}{\partial x} = \lambda \frac{\partial \psi(x)}{\partial x}, \\ \lambda \psi(x) = 0, \\ \lambda \geq 0. \end{cases}$$

Условия дополняющей нежёсткости и неотрицательности.

# Функция Лагранжа для SVM

Функция Лагранжа

$$\mathcal{L}(w, w_0, \xi; \lambda, \eta) = \frac{w^2}{2} + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \lambda_i (y^i (x^i w - w_0) - 1 + \xi_i) - \sum_{i=1}^N \xi_i \eta_i.$$

После преобразований

$$\mathcal{L}(w, w_0, \xi; \lambda, \eta) = \frac{w^2}{2} - \sum_{i=1}^N \lambda_i (y^i (x^i w - w_0) - 1) - \sum_{i=1}^N \xi_i (\lambda_i + \eta_i - C).$$

# Условия стационарности

Дифференцируем

$$\left\{ \begin{array}{ll} \frac{\partial \mathcal{L}}{\partial w} = 0, & w = \sum_{i=1}^N \lambda_i y^i x^i, \\ \frac{\partial \mathcal{L}}{\partial w_0} = 0, & \sum_{i=1}^N \lambda_i y^i = 0, \\ \frac{\partial \mathcal{L}}{\partial \xi_i} = 0, & \eta_i + \lambda_i = C, \quad i = 1, \dots, N. \end{array} \right.$$

## Двойственная задача

Выразим всё через  $\lambda_i$

$$\begin{cases} -\mathcal{L}(\lambda) = \sum_{i=1}^N \lambda_i + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j y^i y^j x^i x^j \rightarrow \min_{\lambda} \\ 0 \leq \lambda_i \leq C, \quad i = 1, \dots, N, \\ \sum_{i=1}^N \lambda_i y^i = 0. \end{cases}$$

Точки, для которых  $\lambda_i > 0$ , называются опорными векторами. Заметим,  $x^i$  что входят только через скалярные произведения.

Итоговый классификатор имеет вид

$$f(x) = \text{sign} \left( \sum_{i=1}^N \lambda_i y^i x^i x - w_0 \right).$$

## Замечания

Оптимизационную задачу можно решать методом активных ограничений (incremental active set method, INCAS), частным случаем которого является симплекс-метод.

Константу  $C$  обычно выбирают по критерию скользящего контроля.

Можно применить фильтрацию выбросов. Сначала задача решается при некотором  $C$ , и из выборки удаляется небольшая доля объектов, имеющих наибольшую величину ошибки  $\xi_i$ . После этого задача решается заново по усечённой выборке.

## Ядра и спрямляющие пространства

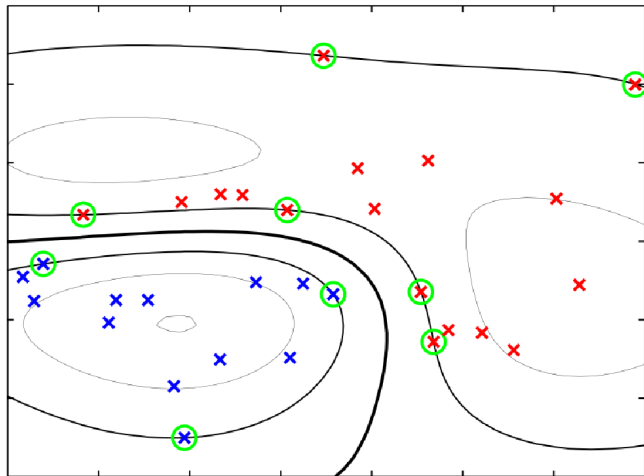
Можно перейти от исходного пространства  $X$  к новому пространству  $H$  с помощью некоторого преобразования  $\psi : X \rightarrow H$ . Пространство  $H$  называют спрямляющим. После этого скалярное произведение примет вид

$$K(x, x') = \psi(x) \cdot \psi(x').$$

Можно вообще не вводить новое пространство, а сразу задать ядро  $K$ .



# Пример работы SVM с ядром



# Открытые проблемы

- Как соотносится SVN с метрическими методами.
- Возможно ли придумать обобщение SVM, включающее дискриминант Фишера как частный случай.
- Является ли разреженность достоинством.
- Как выбирать ядро.
- Низкая эффективность при сильно перекрывающихся распределений классов.