Министерство образования и науки Российской Федерации

Самарский государственный аэрокосмический университет
имени академика С.П. Королева

МЕТОД ОПОРНЫХ ВЕКТОРОВ

Методические указания к лабораторной работе № 4
по курсу «МЕТОДЫ РАСПОЗНАВАНИЯ ОБРАЗОВ»

САМАРА
2015

Цель работы - изучение теоретических основ и экспериментальное исследование метода опорных векторов построения классификаторов для распознавания образов.

1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ЛАБОРАТОРНОЙ РАБОТЫ

1.1. Метод опорных векторов для случая линейно разделимых классов

Рассмотрим *линейный классификатор*, дискриминантная функция которого допускает представление в следующем виде:

$$d(\bar{x}) = \bar{w}^T \bar{x} + w_N \tag{1}$$

где $\bar{x} = (x_0,..., x_{N-1})^T \in D$ - вектор признаков, который определяет образ объекта ω в пространстве признаков *D*, подлежащего классификации, $\bar{w} = (w_0,..., w_{N-1})^T$ - вектор весовых коэффициентов классификатора, w_N - пороговое значение. Процесс принятия решения о номере класса текущего объекта производится в соответствии со следующим правилом:

$$d(\bar{x}) = \sum_{i=0}^{N-1} w_i x_i + w_N \begin{matrix} > \\ < \end{matrix} 0 \quad \Rightarrow \quad \bar{X} \in \begin{cases} D_0 \\ D_1 \end{cases} \tag{2}$$

Здесь D_0 и D_1 - области пространства признаков, соответствующие принятию классификатором решения, о принадлежности объекта-прообраза ω в классе Ω_0 или Ω_1, соответственно. Определим также *переменную правильной классификации* в виде:

$$r(\bar{x}(\omega)) = \begin{cases} 1, & \omega \in \Omega_1, \\ -1, & \omega \in \Omega_0. \end{cases}$$

Рассмотрим линейно разделимую обучающую выборку $\{x_i, r_j\}_{j=0}^{N-1}$ ($r_j \equiv r(\bar{x}_j)$), то есть выборку, для которой найдутся параметры классификатора (1), такие что:

$$\forall j = \overline{0, N-1} \quad \bar{w}^T \bar{x}_j + w_N > 0 \wedge r_j = 1 \quad \vee \quad \bar{w}^T \bar{x}_j + w_N < 0 \wedge r_j = -1.$$

Учитывая, что "масштаб" значений дискриминантной функции не влияет на результаты классификации, можно так определить параметры классификатора, чтобы выполнялись следующие ограничения:

$$\forall j = \overline{0, N-1} \quad \left| \overline{w}^T \overline{x}_j + w_N \right| \geq 1, \tag{3}$$

причем равенство в указанном неравенстве происходит только для ближайших точек к разделяющей гиперплоскости (см. рисунок 1).

Идея *метода опорных векторов* (англ.: SVM - Support Vectors Machine) [Вапник, Воронцов] заключается в поиске вектора коэффициентов $\overline{w}$ и порогового значения w_N, удовлетворяющих условию (3) и обеспечивающих *максимальную ширину разделяющей полосы* (см. рисунок 1):

$$\left| \overline{w}^T \overline{x} + w_N \right| \leq 1. \tag{4}$$

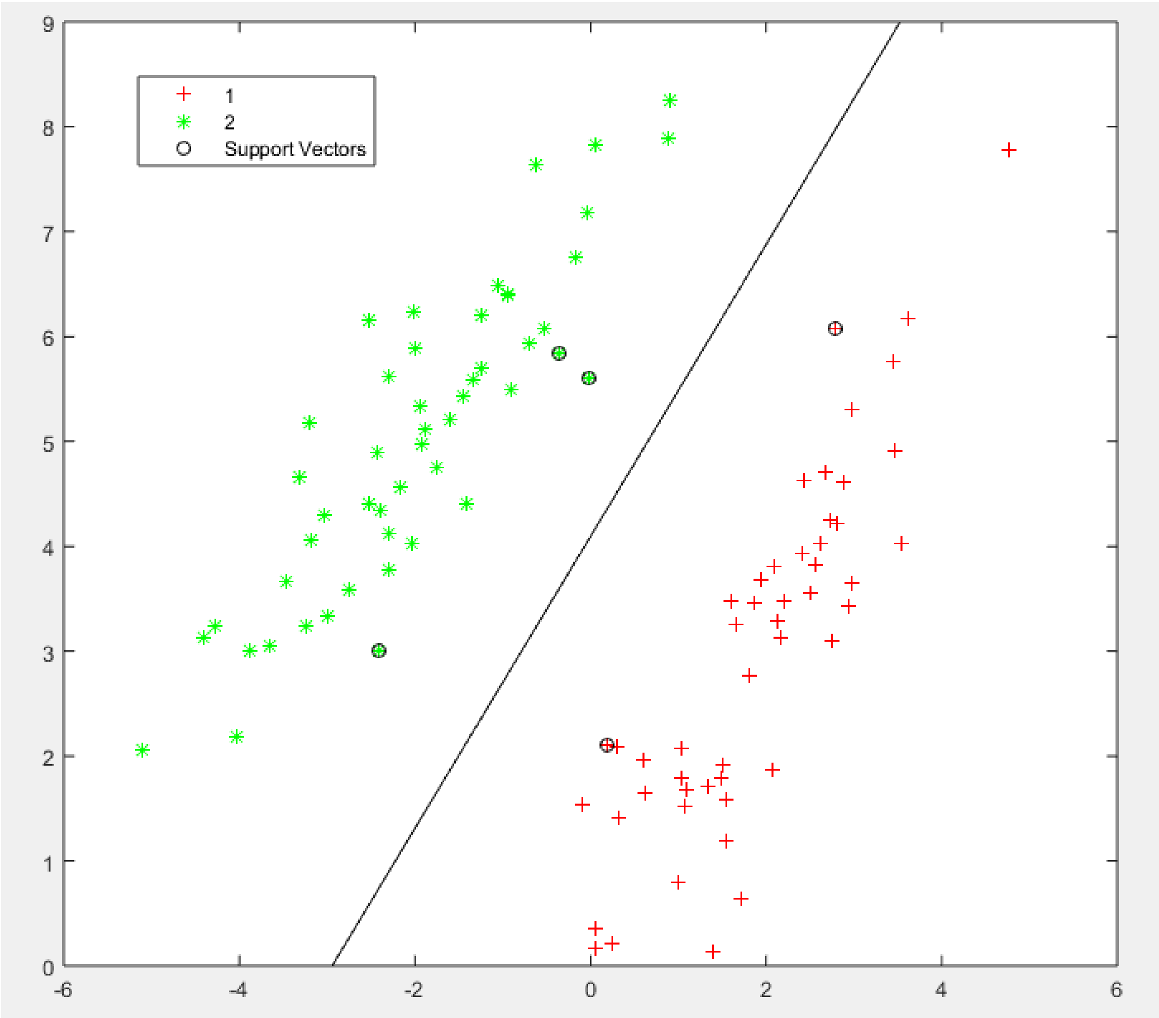

Рис.1. *Разделяющая гиперплоскость и разделяющая полоса для линейно разделимых классов*

Можно показать, что ширина разделяющей полосы в данном случае задается величиной $2\left\| \overline{w} \right\|^{-1}$. Тогда окончательный вид *задачи построения линейного классификатора по методу опорных векторов* следующий:

$$\begin{cases} \overline{w}^T \overline{w} \to \min\limits_{\overline{w}}, \\ \left(\overline{w}^T \overline{x}_j + w_N\right) \cdot r_j \geq 1, \quad j = \overline{0, N-1}. \end{cases} \tag{5}$$

По теореме Куна-Такера [МетодОптим] эта задача квадратичного программирования с линейными ограничениями-неравенствами сводится к задаче поиска седловой точки функции Лагранжа:

$$\begin{cases} \Im\left(\overline{w}, w_N, \overline{\lambda}\right) = \frac{1}{2} \overline{w}^T \overline{w} - \sum\limits_{j=0}^{N-1} \lambda_j \left[r_j \left(\overline{w}^T \overline{x}_j + w_N\right) - 1\right] \to \min\limits_{\overline{w}} \max\limits_{\overline{\lambda}}, \\ \lambda_j \geq 0, \\ \left(\overline{w}^T \overline{x}_j + w_N = r_j\right) \vee \left(\lambda_j = 0\right), \quad j = \overline{0, N-1}. \end{cases} \tag{6}$$

Здесь $\overline{\lambda} = \left(\lambda_0, \ldots, \lambda_{N-1}\right)^T$ - вектор двойственных переменных. Можно показать, что данная задача эквивалентна следующей задаче квадратичного программирования относительно двойственных переменных:

$$\begin{cases} -\sum\limits_{j=0}^{N-1} \lambda_j + \frac{1}{2} \sum\limits_{j=0}^{N-1} \sum\limits_{i=0}^{N-1} \lambda_j \lambda_i r_j r_i \cdot x_j^T x_i \to \min\limits_{\overline{\lambda}}, \\ \sum\limits_{j=0}^{N-1} \lambda_j r_j = 0; \\ \lambda_j \geq 0, \quad j = \overline{0, N-1}. \end{cases} \tag{7}$$

Определение. *Опорным вектором* выборки $\left\{x_j, r_j\right\}_{j=0}^{N-1}$ называется вектор x_{j^*}, для которого выполняется условие:

$$\left(\overline{w}^T \overline{x}_j + w_N = r_j\right) \wedge \left(\lambda_j > 0\right). \tag{8}$$

Имея решение задачи (7) относительно вектора двойственных переменных, мы находим вектор весов классификатора по формуле:

$$\overline{w} = \sum_{j=0}^{N-1} \lambda_j r_j \cdot x_j. \tag{9}$$

Очевидно, что *решение есть линейная комбинация опорных векторов*! Далее имеем:

$$w^T x = \sum_{j=0}^{N-1} \lambda_j r_j \cdot x_j^T x \tag{10}$$

а пороговое значение может быть определено из условия:

$$\bar{w}^T \bar{x}_{j^*} + w_N = r_j ,$$

где $\lambda_j > 0$, то есть по опорным векторам.

1.2. Метод опорных векторов для случая линейно неразделимых классов

Введя дополнительные переменные $s_j \geq 0$, отвечающие за максимально допустимое "нарушение" объектом x_j границы разделяющей полосы, задачу (5) можно переписать в следующем виде:

$$\begin{cases} \frac{1}{2} w^T w + C \sum_{j=0}^{N-1} s_j \to \min_{\bar{w}, \bar{s}}, \\ \left(\bar{w}^T \bar{x}_j + w_N\right) \cdot r_j \geq 1 - s_j, \\ s_j \geq 0 \quad j = \overline{0, N-1}. \end{cases} \tag{11}$$

Здесь величина C - параметр алгоритма, определяющий компромисс между шириной разделяющей полосы и величиной "нарушений" (ошибок), определяемой суммой $\sum_{j=0}^{N-1} s_j$.Проводя рассуждения, аналогичные представленным в п.1, получаем в итоге следующую задачу квадратичного программирования с двойственными переменными:

$$\begin{cases} -\sum_{j=0}^{N-1} \lambda_j + \frac{1}{2} \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} \lambda_j \lambda_i r_j r_i \cdot x_j^T x_i \to \min_{\bar{\lambda}}, \\ \sum_{j=0}^{N-1} \lambda_j r_j = 0; \\ 0 \leq \lambda_j \leq C, \quad j = \overline{0, N-1}. \end{cases} \tag{12}$$

Данная задача отличается от задачи (7) только допустимым диапазоном двойственных переменных. Определение величины C осуществляется экспериментальным путем.

1.3 Использование ядер в методе опорных векторов и переход в пространства высокой размерности

Пусть определено некоторое отображение $\psi : D \to H$, где H - гильбертово пространство (пространство с определенным в нем скалярным произведением). Используем для классификации образов вместо исходных описаний x образы этих векторов признаков в новом пространстве, то есть вектора вида $\psi(x)$. Тогда задача (12) будет иметь следующий вид:

$$\begin{cases} -\sum_{j=0}^{N-1}\lambda_j + \frac{1}{2}\sum_{j=0}^{N-1}\sum_{i=0}^{N-1}\lambda_j\lambda_i r_j r_i \cdot \psi(x_j)^T\psi(x_i) \to \min_{\overline{\lambda}}, \\ \sum_{j=0}^{N-1}\lambda_j r_j = 0; \\ 0 \le \lambda_j \le C, \quad j = \overline{0, N-1}. \end{cases} \tag{13}$$

Здесь $\psi(x_j)^T\psi(x_i)$ - скалярное произведение в новом гильбертовом пространстве описаний, то есть симметричная и положительно определенная функций. Обозначим $K(x_j, x_i) \equiv \psi(x_j)^T\psi(x_i)$, такие функции называют *ядрами*. Тогда задача (13) может быть переписана в виде:

$$\begin{cases} -\sum_{j=0}^{N-1}\lambda_j + \frac{1}{2}\sum_{j=0}^{N-1}\sum_{i=0}^{N-1}\lambda_j\lambda_i r_j r_i \cdot K(x_j, x_i) \to \min_{\overline{\lambda}}, \\ \sum_{j=0}^{N-1}\lambda_j r_j = 0; \\ 0 \le \lambda_j \le C, \quad j = \overline{0, N-1}. \end{cases} \tag{14}$$

Данная задача может быть решена так же, как и задачи (12)-(13), при этом *информация о преобразовании вектора признаков* $\psi(x)$ *в явном виде не требуется* - достаточно знать только вид ядра $K(\cdot,\cdot)$.

Имея решение задачи (14), классификатор можно определить на основании следующих соотношений. Выражение для классификатора (10) преобразуется к виду:

$$\overline{w}^T\overline{x} = \sum_{j=0}^{N-1}\lambda_j r_j \cdot K(\overline{x}_j, \overline{x}) \tag{15}$$

и очевидным образом вычисляется с учетом известного ядра и найденного вектора двойственных переменных $\overline{\lambda}$. Пороговое значение рассчитывается по тем же правилам, что и в п.1, что с учетом (15) дает выражение:

$$w_N = r_j - \sum_{j=0}^{N-1} \lambda_j r_j \cdot K\left(x_j, \bar{x}_{j^*}\right).$$

Примеры некоторых широко используемых ядер представлены в следующей таблице. Заметим, что полиномиальное однородное ядро для d=2 эквивалентно использованию отображения $\psi(x)$, формирующего новые признаки вида: $\{x_i x_j\}_{0 \le i \le j < n}$ [Вор].

Таблица. *Примеры ядер*

Наименование	Вид ядра
Полиномиальное однородное	$K(x, y) = \left(x^T y\right)^d$
Полиномиальное неоднородное	$K(x, y) = \left(x^T y + 1\right)^d$
Радиальная функция	$K(x, y) = \exp\left(-\gamma \lVert x - y \rVert^2\right), \quad \gamma > 0$
Радиальная функция Гаусса	$K(x, y) = \exp\left(-\frac{\lVert x - y \rVert^2}{2\sigma^2}\right)$
Сигмоидальная функция	$K(x, y) = \tanh\left(\gamma x^T y + c\right), \quad \gamma > 0, c < 0$

1.4 Достоинства и недостатки метода опорных векторов

Достоинства метода опорных векторов

1. Высокая "надежность" (обобщающая способность) решающего правила.

2. Универсальность, отсутствие формальных ограничений по использованию алгоритма.

2. Возможность неявного выбора пространства классификации путем явного выбора ядер.

4. Возможность неявного перехода (с использованием ядер) в пространство более высокой размерности, где классы оказываются разделимы с большей "вероятностью".

Недостатки метода опорных векторов

1. Сильное влияние шумов, так как классификатор напрямую зависит от выбранных опорных векторов.

2. Необходимость выбора параметра C для случая отсутствия линейной разделимости классов.

3. Отсутствие регулярного метода подбора ядра, связанного с выбором целевого пространства классификации.

4. Вычислительная сложность решения задач квадратичного программирования (7), (12), (14), которая делает затруднительным получения точного решения для больших обучающих выборок.