Метод опорных признаков в задаче распознавания объектов двух классов

Н. А. Савинов

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Научный руководитель д.т.н., в.н.с. ВЦ РАН В. В. Моттль

Москва,
2011 г.

План презентации

- Постановка задачи для метода опорных векторов
- Свойства решения для метода опорных векторов
- Проблема переобучения
- Метод опорных признаков
- Способы решения оптимизационной задачи
- Вычислительный эксперимент

Постановка задачи для классического метода опорных векторов

- Задана обучающая выборка $(\mathbf{x}_j, y_j),\ j = 1, ..., N$, где $\mathbf{x_j} \in \mathbb{R}^n$ — признаковые описания объектов, $y_j \in \{-1, 1\}$ — ответы.
- Рассматривается класс линейных решающих правил: $y = sign(\sum_{i=1}^{n} a_i x_i + b)$
- Выбор оптимального правила сводится к настройке параметров $\mathbf{a}$, b по обучающей выборке на основе критерия:

$$\begin{cases} \frac{1}{2}\mathbf{a}^T\mathbf{a} + c\sum_{j=1}^{N}\delta_j \to min(\mathbf{a}, b, \delta_j), \\ y_j(\mathbf{a}^T\mathbf{x}_j + b) \geq 1 - \delta_j,\ j = 1, ..., N, \\ \delta_j \geq 0,\ j = 1, ..., N. \end{cases}$$

Двойственная задача

Пусть $\lambda_j,\ j = 1, ..., N$ — двойственные переменные. Тогда двойственная задача имеет вид:

$$
\begin{cases}
\sum\limits_{j=1}^{N} \lambda_j - \frac{1}{2} \sum\limits_{i=1}^{N} \sum\limits_{j=1}^{N} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \to max(\lambda_j), \\
0 \le \lambda_j \le c,\ j = 1, ..., N, \\
\sum\limits_{j=1}^{N} y_j \lambda_j = 0.
\end{cases}
$$

Свойства двойственной задачи оптимизации:

- Квадратичная
- Вогнутая

Данная задача эффективно решается методом Sequential Minimal Optimization за $O(N^2)$ действий (Platt, 2001).

Решение исходной задачи

Решение прямой задачи выражается через решение двойственной с использованием условий Каруша-Куна-Таккера:

- $\mathbf{a} = \sum_{j=1}^{N} \lambda_j y_j \mathbf{x}_j$
- $b = med\{y_j - \mathbf{a}^T \mathbf{x}_j,\ \lambda_j > 0,\ j = 1, ..., N\}$

Свойства решения

- Направляющий вектор **a** разделяющей гиперплоскости является линейной комбинацией признаковых описаний объектов обучения.
- Объекты, для которых $\lambda_j = 0$, не влияют на положение разделяющей гиперплоскости.
- В случае линейной разделимости выборки метод максимизирует ширину разделяющей полосы между классами.

Переобучение в методе опорных векторов

- В случае, когда $n \geq N$ (то есть когда размерность признакового описания велика, а объектов в обучении мало), метод опорных векторов склонен к переобучению.
- Причина: l_2-регуляризатор $\frac{1}{2}\mathbf{a}^T\mathbf{a}$, введенный для улучшения обобщающей способности, не отбирает признаки.

Предлагается: l_2-l_1 регуляризация

- Вводится регуляризатор:
  $\sum_{i=1}^{n}((1-\mu)a_i^2+\mu|a_i|)$, где $\mu\in[0,1)$ — параметр,
  позволяющий регулировать селективность метода.
- Получаем критерий обучения:

$$\begin{cases} \sum_{i=1}^{n}((1-\mu)a_i^2+\mu|a_i|)+c\sum_{j=1}^{N}\delta_j \to min(\mathbf{a},b,\delta_j),\\ y_j(\mathbf{a}^T\mathbf{x}_j+b)\geq 1-\delta_j,\ j=1,...,N,\\ \delta_j\geq 0,\ j=1,...,N. \end{cases}$$

Двойственная задача

$$\begin{cases} \sum\limits_{j=1}^{N} \lambda_j - \frac{1}{4(1-\mu)} \sum\limits_{i=1}^{n} [min\{\mu + \sum\limits_{j=1}^{N} y_j \lambda_j x_{ij}, 0, \mu - \sum\limits_{j=1}^{N} y_j \lambda_j x_{ij}\}]^2 \to max(\lambda_j), \\ 0 \le \lambda_j \le c, \ j = 1, ..., N, \\ \sum\limits_{j=1}^{N} y_j \lambda_j = 0. \end{cases}$$

Свойства:

- Неквадратичная
- Вогнутая

Решение существует, каждое локальное решение является глобальным.

Решение исходной задачи

Полностью определяется через решение двойственной задачи:

$$a_i = \begin{cases} \frac{\sum\limits_{j=1}^{N} y_j \lambda_j x_{ij} - \mu}{2(1-\mu)}, \text{если } \sum\limits_{j=1}^{N} y_j \lambda_j x_{ij} > \mu, \\ \frac{\sum\limits_{j=1}^{N} y_j \lambda_j x_{ij} + \mu}{2(1-\mu)}, \text{если } \sum\limits_{j=1}^{N} y_j \lambda_j x_{ij} < -\mu, \\ 0, \text{если } \left|\sum\limits_{j=1}^{N} y_j \lambda_j x_{ij}\right| \leq \mu, \end{cases}$$

$$b = -\frac{\sum\limits_{0<\lambda_j<c} \lambda_j \mathbf{a}^T \mathbf{x}_j + c \sum\limits_{\lambda_j = c} y_j}{\sum\limits_{0<\lambda_j<c} \lambda_j}.$$

Свойства решения

- Некоторые a_i оказываются равными 0, то есть производится отбор признаков.
- Поэтому предложенный метода был назван методом опорных признаков.

Способы решения двойственной задачи

Предлагаются 2 способа:

- Сведение к квадратичной.
- Решение напрямую с помощью обобщения метода Sequential Minimal Optitimization на неквадратичный случай.

Квадратичный вид двойственной задачи

$$
\begin{cases}
\sum\limits_{j=1}^{N} \lambda_j - \frac{1}{4(1-\mu)} \sum\limits_{i=1}^{n} \gamma_i^2 \to max(\lambda_j, \gamma_i), \\
|\sum\limits_{j=1}^{N} \lambda_j y_j x_{ij}| \leq \gamma_i + \mu, i = 1, ..., N, \\
0 \leq \lambda_j \leq c, j = 1, ..., N, \\
\sum\limits_{j=1}^{N} y_j \lambda_j = 0.
\end{cases}
$$

- Далее можно использовать стандартные программные пакеты решения задач квадратичного программирования, требующие $O(N^3)$ действий.
- Недостаток: введение дополнительных переменных.

Обобщение метода SMO

- Для использования метода SMO достаточно решить двойственную задачу при 2 свободных переменных λ_{k_1}, λ_{k_2} и фиксированных остальных.
- В силу наличия ограничения-равенства на переменные задача сводится к одномерной максимизации по λ_{k_1}.
- Целевая функция дифференцируема по λ_{k_1}:

$$\frac{dW}{d\lambda_{k_1}} = (1 - y_{k_1} y_{k_2}) - \frac{y_{k_1}}{2(1-\mu)} (S_1 + S_2)\Big|_{\lambda_{k_2} = \gamma - y_{k_1} y_{k_2} \lambda_{k1}}, \text{ где}$$

$$S_1 = \sum_{i:\, \mathbf{X}_{yi}\lambda < -\mu} (\mathbf{X}_{yi}\lambda + \mu)(x_{ik_1} - x_{ik_2}),$$

$$S_2 = -\sum_{i:\, \mathbf{X}_{yi}\lambda > \mu} (-\mathbf{X}_{yi}\lambda + \mu)(x_{ik_1} - x_{ik_2}),$$

$\mathbf{X}_{yi}$ — вектор-строка матрицы $(y_j x_{ij})$.

Обобщение метода SMO

- Производная представляет собой монотонную кусочно-линейную функцию, состоящую из $(2n+1)$ участков.
- Для быстрого поиска участка, на котором достигается 0, испозуется метод дихотомии.
- По найденным концам вычисляются параметры линейного участка, находится точка 0 производной.
- Берется проекция найденной точки безусловного максимума на множество, задаваемое ограничениями-неравенствами. Таким образом, получаем решение задачи условной максимизации при 2 свободных переменных.

Эксперимент на модельных данных

Описание данных:

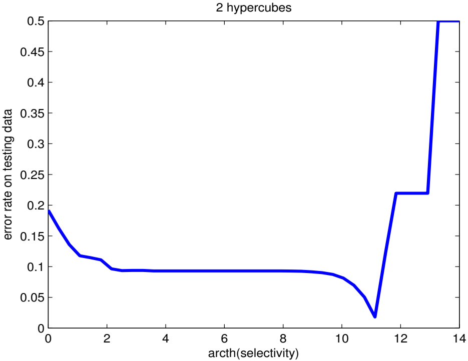
- $n = 50$ — число признаков, $N_1 = 25$ — число объектов первого класса, $N_2 = 25$ — число объектов второго класса.
- Данные генерировались с равномерным распределением внутри двух гиперкубов, касающихся по одной из граней. Нормаль к этой грани направлена по вектору $(5, 4, 3, 2, 1, 0, \dots, 0)$.
- Таким образом, число реальных признаков, по которым выборка разделима, равно 5.

Обобщающая способность контролировалась двумя способами:

- На контрольной выборке из $N = 100000$ объектов, которая генерировалась тем же способом.
- С помощью Cross Validation.

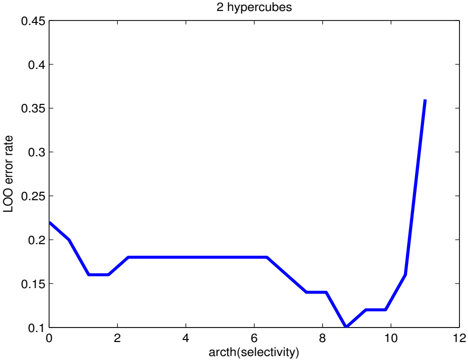
Результат на контрольной выборке

- В точке минимума ошибки алгоритм выделяет 4 признака из 5 правильных.
- Для целей масштабирования по оси X отложена не селективность μ, а величина $arcth(\mu)$.



Результат с использованием Cross Validation

- В точке минимума ошибки алгоритм выделяет 13 признаков, содержащих 5 правильных.
- Отбрасываются 37 шумовых признаков.



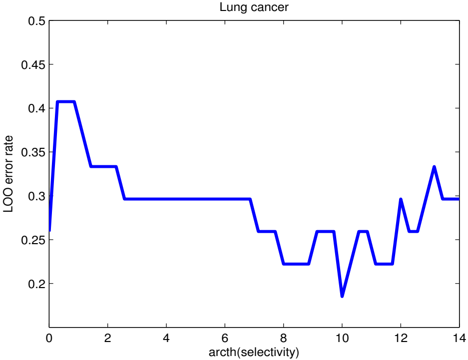
Эксперимент на реальных данных: Lung Cancer

Описание данных:

- $n = 55$ - число признаков, $N_1 = 9$ - число больных пациентов, $N_2 = 18$ - число здоровых пациентов.
- Источник данных: интернет-репозиторий задач машинного обучения UCI.

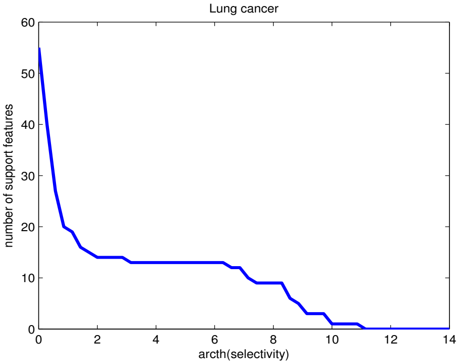
Результат с использованием Cross Validation

В точке минимума ошибки алгоритм выделяет 5 признаков.



Убывание числа признаков с увеличением селективности

При приближении селективности к 1 алгоритм отбрасывает все признаки.



Результаты

- Предложен регуляризатор l_2-l_1 с параметром селективности.
- Получена двойственная задача и выражение для решения исходной задачи через решение двойственной.
- Предложено два метода решения двойственной задачи: сведением к квадратичной задаче и с помощью обобщения метода SMO.
- Исследовано поведение метода на реальных и модельных данных с выбором селективности по Cross Validation.