

Лекция 3. Метод опорных векторов

Д. П. Ветров¹ Ю. И. Журавлев¹

¹МГУ, ВМиК, каф. ММП

Курс «Математические основы теории
прогнозирования»

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно separable данных

Случай линейно non-separable данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Задача условной оптимизации

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

Пусть $f(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$ — гладкая функция. Предположим, что нам необходимо найти ее экстремум:

$$f(\mathbf{x}) \rightarrow \underset{\mathbf{x}}{\text{extr}}$$

Для того, чтобы найти экстремум (решить задачу безусловной оптимизации), достаточно проверить условие стационарности:

$$\nabla f(\mathbf{x}) = 0$$

Предположим, что нам необходимо найти экстремум функции при ограничениях:

$$f(\mathbf{x}) \rightarrow \underset{\mathbf{x}}{\text{extr}}$$

$$g(\mathbf{x}) = 0$$

Поверхность ограничения

Лекция 3. Метод опорных векторов

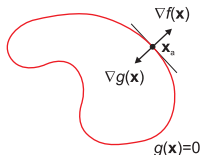
Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознающее распознавание образов



Заметим, что $\nabla g(\mathbf{x})$ ортогонален поверхности ограничения $g(\mathbf{x}) = 0$. Пусть \mathbf{x} и $\mathbf{x} + \boldsymbol{\varepsilon}$ — две близкие точки поверхности. Тогда

$$g(\mathbf{x} + \boldsymbol{\varepsilon}) \simeq g(\mathbf{x}) + \boldsymbol{\varepsilon}^T \nabla g(\mathbf{x})$$

Т.к. $g(\mathbf{x} + \boldsymbol{\varepsilon}) = g(\mathbf{x})$, то $\boldsymbol{\varepsilon}^T \nabla g(\mathbf{x}) \simeq 0$. При стремлении $\|\boldsymbol{\varepsilon}\| \rightarrow 0$ получаем $\boldsymbol{\varepsilon}^T \nabla g(\mathbf{x}) = 0$. Т.к. $\boldsymbol{\varepsilon}$ параллелен поверхности $g(\mathbf{x}) = 0$, то $\nabla g(\mathbf{x})$ является нормалью к этой поверхности.

Функция Лагранжа

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Необходимым условием оптимальности является ортогональность $\nabla f(\mathbf{x})$ поверхности ограничения, т.е.:

$$\nabla f + \lambda \nabla g = 0 \quad (1)$$

Здесь $\lambda \neq 0$ — коэффициент Лагранжа. Он может быть любого знака.

Функция Лагранжа

$$L(\mathbf{x}, \lambda) \triangleq f(\mathbf{x}) + \lambda g(\mathbf{x})$$

Тогда

$$\nabla_{\mathbf{x}} L = 0 \quad \Rightarrow \quad \text{условие (1)}$$

$$\frac{\partial}{\partial \lambda} L = 0 \quad \Rightarrow \quad g(\mathbf{x}) = 0$$

Функция Лагранжа. Пример.

Лекция 3. Метод
опорных
векторов

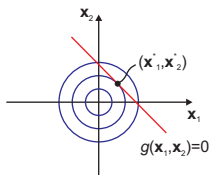
Ветров,
Журавлев

Ликбез

Метод опорных
векторов для
задачи
классификации,
[2]

Метод опорных
векторов для
задачи
регрессии, [2]

Беспознаковое
распознавание
образов



$$f(x_1, x_2) = 1 - x_1^2 - x_2^2 \rightarrow \max_{x_1, x_2}$$

$$g(x_1, x_2) = x_1 + x_2 - 1 = 0$$

Функция Лагранжа:

$$L(\mathbf{x}, \lambda) = 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)$$

Условия стационарности:

$$-2x_1 + \lambda = 0$$

$$-2x_2 + \lambda = 0$$

$$x_1 + x_2 - 1 = 0$$

Решение: $(x_1^*, x_2^*) = (\frac{1}{2}, \frac{1}{2})$, $\lambda = 1$.

Ограничение в виде неравенства

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

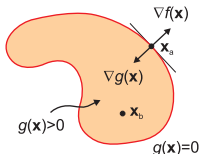
Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

$$f(\mathbf{x}) \rightarrow \max_{\mathbf{x}}$$

$$g(\mathbf{x}) \geq 0$$



Решение	Ограничение	Условие стационарности
Внутри области $g(\mathbf{x}) > 0$	неактивно	$\nabla f(\mathbf{x}) = 0, \nabla_{\mathbf{x}} L = 0, \lambda = 0$
На границе $g(\mathbf{x}) = 0$	активно	$\nabla f(\mathbf{x}) = -\lambda \nabla g(\mathbf{x}),$ $\nabla_{\mathbf{x}, \lambda} L = 0, \lambda > 0$

Условие дополняющей нежесткости:

$$\lambda g(\mathbf{x}) = 0$$

Теорема Каруша-Куна-Таккера

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознаваемое распознавание образов

Пусть $f_i : X \rightarrow \mathbb{R}$, $i = 0, 1, \dots, m$ — выпуклые функции, отображающие нормированное пространство X в прямую, $A \in X$ — выпуклое множество. Рассмотрим следующую задачу оптимизации:

$$f_0(\mathbf{x}) \rightarrow \min; \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \quad \mathbf{x} \in A \quad (2)$$

Теорема

1. Если $\hat{\mathbf{x}} \in \text{absmin}(2)$ — решение задачи, то найдется ненулевой вектор множителей Лагранжа $\boldsymbol{\lambda} \in \mathbb{R}^{m+1}$ такой, что для функции Лагранжа $L(\mathbf{x}) = \sum_{i=0}^m \lambda_i f_i(\mathbf{x})$ выполняются условия:
 - a) стационарности $\min_{\mathbf{x} \in A} L(\mathbf{x}) = L(\hat{\mathbf{x}})$
 - b) дополняющей нежесткости $\lambda_i f_i(\hat{\mathbf{x}}) = 0$, $i = 1, \dots, m$
 - c) неотрицательности $\lambda_i \geq 0$
2. Если для допустимой точки $\hat{\mathbf{x}}$ выполняются условия а)–с) и $\lambda_0 \neq 0$, то $\hat{\mathbf{x}} \in \text{absmin}(2)$
3. Если для допустимой точки $\hat{\mathbf{x}}$ выполняются условия а)–с) и $\exists \bar{\mathbf{x}} \in A : f_i(\bar{\mathbf{x}}) < 0, i = 0, \dots, m$ (условие Слейтера), то $\hat{\mathbf{x}} \in \text{absmin}(2)$

Задача классификации

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

- Рассматривается задача классификации на два класса. Имеется обучающая выборка $(X, t) = \{\mathbf{x}_i, t_i\}_{i=1}^n$, где $\mathbf{x} \in \mathbb{R}^d$, а метка класса $t \in \mathcal{T} = \{-1, 1\}$.
- Необходимо с использованием обучающей выборки построить отображение $A : \mathbb{R}^d \rightarrow \mathcal{T}$, которое для каждого нового входного объекта \mathbf{x}_* выдает его метку класса t_* .

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Метод потенциальных функций, [1]

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

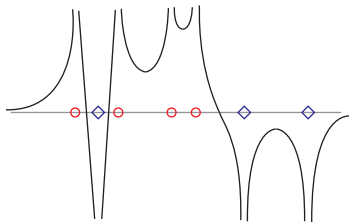
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознающее распознавание образов



В каждом объекте \mathbf{x}_i помещен электрический заряд $t_i q_i$. В качестве разделяющей функции используется потенциал создаваемого поля:

$$f(\mathbf{x}) = \sum_{i=1}^n t_i q_i K(\mathbf{x}, \mathbf{x}_i)$$

Алгоритм обучения:

$$f^{new}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) + K(\mathbf{x}, \mathbf{x}_k), & \text{если } t_k = 1 \text{ и } f(\mathbf{x}_k) \leq 0 \\ f(\mathbf{x}) - K(\mathbf{x}, \mathbf{x}_k), & \text{если } t_k = -1 \text{ и } f(\mathbf{x}_k) \geq 0 \\ f(\mathbf{x}), & \text{иначе} \end{cases}$$

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Разделяющая гиперплоскость

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

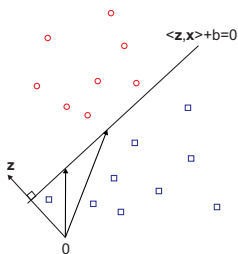
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов



- Гиперплоскость задается направляющим вектором гиперплоскости \mathbf{z} и величиной сдвига b :

$$\{\mathbf{x} \in \mathbb{R}^d \mid \langle \mathbf{z}, \mathbf{x} \rangle + b = 0\}, \quad \mathbf{z} \in \mathbb{R}^d, \quad b \in \mathbb{R}$$

- Если вектор \mathbf{z} имеет единичную длину, то величина $\langle \mathbf{z}, \mathbf{x} \rangle$ определяет длину проекции вектора \mathbf{x} на направляющий вектор \mathbf{z} . В случае произвольной длины скалярное произведение нормируется на $\|\mathbf{z}\|$.

Каноническая гиперплоскость

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

- Если величину направляющего вектора \mathbf{z} и величину сдвига b умножить на одно и то же число, то соответствующая им гиперплоскость не изменится.
- Пара (\mathbf{z}, b) задает **каноническую гиперплоскость** для набора объектов $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$, если

$$\min_{i=1, \dots, n} |\langle \mathbf{z}, \mathbf{x}_i \rangle + b| = 1 \quad (3)$$

- Условие (3) означает, что ближайший вектор к канонической гиперплоскости находится от нее на расстоянии $1/\|\mathbf{z}\|$:

$$\mathbf{x}_i : |\langle \mathbf{z}, \mathbf{x}_i \rangle + b| = 1, \quad \mathbf{x} : \langle \mathbf{z}, \mathbf{x} \rangle + b = 0$$

$$|\langle \mathbf{z}, \mathbf{x}_i - \mathbf{x} \rangle| = 1 \Rightarrow \left| \left\langle \frac{\mathbf{z}}{\|\mathbf{z}\|}, \mathbf{x}_i - \mathbf{x} \right\rangle \right| = \frac{1}{\|\mathbf{z}\|}$$

- Каноническая гиперплоскость определена однозначно с точностью до знака \mathbf{z} и b

Классификатор

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

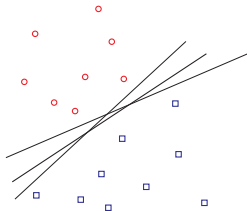
Беспризнаковое распознавание образов

- Будем искать классификатор в виде разделяющей гиперплоскости:

$$\hat{t}(\mathbf{x}) = \text{sign}(y(\mathbf{x})) = \text{sign}(\langle \mathbf{z}, \mathbf{x} \rangle + b)$$

- Предположим, что исходные данные являются линейно разделимыми, т.е.

$$\exists(\mathbf{z}, b) : \hat{t}(\mathbf{x}_i) = t_i \quad \forall i = 1, \dots, n$$



Оптимальная разделяющая гиперплоскость

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

- Зазором гиперплоскости данной точки (\mathbf{x}, t) называется величина:

$$\rho_{(z,b)}(\mathbf{x}, t) = \frac{t(\langle \mathbf{z}, \mathbf{x} \rangle + b)}{\|\mathbf{z}\|}$$

Зазором гиперплоскости называется величина:

$$\rho_{(z,b)} = \min_{i=1, \dots, n} \rho_{(z,b)}(\mathbf{x}_i, t_i)$$

- Для корректно распознаваемого объекта его величина зазора соответствует расстоянию от этого объекта до гиперплоскости. Отрицательная величина зазора соответствует об ошибочной классификации объекта.
- **Оптимальная разделяющая гиперплоскость — гиперплоскость с максимальной величиной зазора**

Оптимальная разделяющая гиперплоскость.

Пример 1.

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

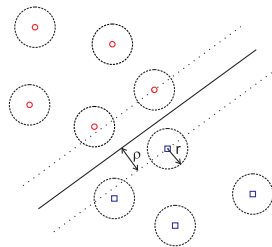
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов



- Предположим, что все объекты тестовой выборки получены путем небольших смещений относительно обучающих объектов, т.е. для объекта обучения (x, t) тестовый объект может быть представлен как $(x + \Delta x, t)$, причем $\|\Delta x\| \leq r$.
- Гиперплоскость с величиной зазора $\rho > r$ корректно классифицирует выборку.

Оптимальная разделяющая гиперплоскость.

Пример 2.

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

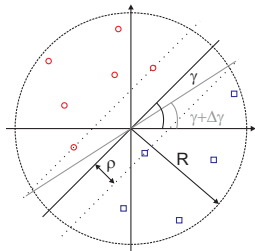
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов



Если объекты располагаются на достаточном расстоянии от гиперплоскости, то небольшое изменение параметров (z, b) не меняет корректного разделения данных.

Оптимальная разделяющая гиперплоскость. Стат. теория Вапника-Червоненкиса, [2]

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Веспризнаковое распознавание образов

- Пусть имеется некоторое объективное распределение $p(\mathbf{x}, t)$. Обучающая и тестовая совокупности являются н.о.р. выборками из этого распределения.
- Пусть имеется семейство классификаторов $\{f(\mathbf{x}, \mathbf{w}) : \mathbb{R}^d \rightarrow \mathcal{T} | \mathbf{w} \in \Omega\}$ и функция потерь $L : \mathcal{T} \times \mathcal{T} \rightarrow \mathbb{R}_+$.
- **Средним риском** называется мат.ожидание функции потерь:

$$R(\mathbf{w}) = \mathbb{E}_{p(\mathbf{x}, t)} L(\cdot) = \int L(t, f(\mathbf{x}, \mathbf{w})) p(\mathbf{x}, t) d\mathbf{x} dt$$

- **Эмпирическим риском** называется следующая величина:

$$R(\mathbf{w}) = \frac{1}{n} \sum_{i=1}^n L(t_i, f(\mathbf{x}_i, \mathbf{w}))$$

Оптимальная разделяющая гиперплоскость. Стат. теория Валника-Червоненкиса, [2]

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов

Теорема (Vapnik, 1995)

С вероятностью $0 < \eta \leq 1$ выполняется следующее неравенство:

$$R(\mathbf{w}) \leq R(\mathbf{w}) + \sqrt{\frac{h(\ln(2m/h) + 1) - \ln(\eta/4)}{m}}$$

Здесь h — положительная величина, называемая ВЧ-размерностью.

Теорема (Vapnik, 1995)

Допустим, что вектора $\mathbf{x} \in \mathbb{R}^d$ принадлежат сфере радиуса R . Тогда для семейства разделяющих гиперплоскостей с величиной зазора ρ верно следующее:

$$h \leq \min \left(\left\lceil \frac{R^2}{\rho^2} \right\rceil, d \right) + 1$$

Постановка задачи оптимизации

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциалных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

- Предположим, что в выборке присутствуют объекты двух классов, т.е. $\exists i, j : t_i = 1, t_j = -1$.
- Для построения канонической гиперплоскости с максимальной величиной зазора, корректно разделяющей данные, необходимо решить следующую задачу оптимизации:

$$\frac{1}{2} \|z\|^2 \rightarrow \min_{z, b}$$

$$t_i(\langle z, x_i \rangle + b) \geq 1, \forall i = 1, \dots, n$$

Функция Лагранжа

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознающее распознавание образов

Функция Лагранжа

$$L(\mathbf{z}, b, \mathbf{w}) = \frac{1}{2} \|\mathbf{z}\|^2 - \sum_{i=1}^n w_i (t_i (\langle \mathbf{z}, \mathbf{x}_i \rangle + b) - 1) \rightarrow \min_{\mathbf{z}, b} \max_{\mathbf{w}}$$

Коэффициенты Лагранжа $w_i \geq 0, \forall i = 1, \dots, n$.

$$\frac{\partial}{\partial \mathbf{z}} L(\mathbf{z}, b, \mathbf{w}) = 0 \Rightarrow \mathbf{z}_* = \sum_{i=1}^n w_i t_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b} L(\mathbf{z}, b, \mathbf{w}) = 0 \Rightarrow \sum_{i=1}^n w_i t_i = 0$$

$$\begin{aligned} L(\mathbf{z}_*, b_*, \mathbf{w}) &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j t_i t_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^n \sum_{j=1}^n w_i w_j t_i t_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \\ &+ \sum_{i=1}^n w_i = \sum_{i=1}^n w_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j t_i t_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \rightarrow \max_{\mathbf{w}} \end{aligned}$$

Двойственная задача оптимизации

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов

$$\sum_{i=1}^n w_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j t_i t_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \rightarrow \max_{\mathbf{w}}$$

$$\sum_{i=1}^n t_i w_i = 0$$

$$w_i \geq 0, \quad \forall i = 1, \dots, n$$

Решение

$$\hat{t}(\mathbf{x}) = \text{sign}(\langle \mathbf{z}_*, \mathbf{x} \rangle + b_*) = \text{sign} \left(\sum_{i=1}^n w_i^* t_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b_* \right)$$

Опорные вектора

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

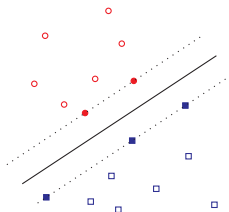
Метод опорных векторов для задачи регрессии, [2]

Беспознаваемое распознавание образов

Условие дополняющей нежесткости:

$$w_i^*(t_i(\langle \mathbf{z}_*, \mathbf{x}_i \rangle + b_*) - 1) = 0 \quad (4)$$

Из (4) следует, что объекты обучающей выборки, для которых $w_i^* > 0$, лежат точно на границе гиперплоскости. Они называются **опорными векторами**.



Значение b_* может быть получено из (4) для любого опорного вектора. При этом с вычислительной точки зрения более устойчивой процедурой является усреднение по всем таким объектам.

Разреженность решения

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов

Решающее правило

$$\hat{t}(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^n w_i t_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b \right)$$

В решающее правило входят только те объекты обучения, для которых $w_i > 0$ (опорные векторы). Такое правило называется **разреженным (sparse model)**. Разреженные модели обладают высокой скоростью распознавания в больших объемах данных, а также «проливают свет» на структуру обучающей совокупности, выделяя наиболее релевантные с точки зрения классификации объекты.

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Ослабляющие коэффициенты

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов

На практике данные не являются, как правило, линейно разделимыми. Кроме того, даже линейно разделимые выборки могут содержать помехи, ошибочные метки классов и проч. Практический метод распознавания должен учитывать подобные ситуации.

Предположим, что $\forall(\mathbf{z}, b) \exists \mathbf{x}_i : \rho_{(\mathbf{z}, b)}(\mathbf{x}_i, t_i) < 0$. Позволим некоторым из ограничений не выполняться путем введения ослабляющих коэффициентов:

$$t_i(\langle \mathbf{z}, \mathbf{x}_i \rangle + b) \geq 1 \quad \forall i = 1, \dots, n \longrightarrow \begin{cases} t_i(\langle \mathbf{z}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \\ \xi_i \geq 0, \quad \forall i = 1, \dots, n \end{cases}$$

При этом потребуем, чтобы количество нарушений (количество ошибок на обучении) было бы как можно меньшим:

$$\frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min_{\mathbf{z}, b, \xi}$$

Постановка задачи оптимизации

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Веспризнаковое распознавание образов

$$\frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min_{\mathbf{z}, b}$$

$$t_i(\langle \mathbf{z}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \quad \forall i = 1, \dots, n$$

$$\xi_i \geq 0$$

Здесь $C \geq 0$ — некоторый действительный параметр, играющий роль параметра регуляризации

Функция Лагранжа

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Веспризнаковое распознавание образов

Функция Лагранжа

$$L(\mathbf{z}, b, \boldsymbol{\xi}, \mathbf{w}, \mathbf{v}) = \frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n w_i [t_i (\langle \mathbf{z}, \mathbf{x}_i \rangle + b) - 1 + \xi_i] - \sum_{i=1}^n v_i \xi_i \rightarrow \min_{\mathbf{z}, b, \boldsymbol{\xi}} \max_{\mathbf{w}, \mathbf{v}}$$

Коэффициенты Лагранжа $w_i \geq 0$, $v_i \geq 0$.

$$\frac{\partial}{\partial \mathbf{z}} L(\mathbf{z}, b, \boldsymbol{\xi}, \mathbf{w}, \mathbf{v}) = 0 \quad \Rightarrow \quad \mathbf{z}_* = \sum_{i=1}^n w_i t_i \mathbf{x}_i$$

$$\frac{\partial}{\partial b} L(\mathbf{z}, b, \boldsymbol{\xi}, \mathbf{w}, \mathbf{v}) = 0 \quad \Rightarrow \quad \sum_{i=1}^n w_i t_i = 0$$

$$\frac{\partial}{\partial \xi_i} L(\mathbf{z}, b, \boldsymbol{\xi}, \mathbf{w}, \mathbf{v}) = 0 \quad \Rightarrow \quad w_i + v_i = C$$

Двойственная задача оптимизации

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

$$\sum_{i=1}^n w_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_i w_j t_i t_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \rightarrow \max_{\mathbf{w}}$$

$$\sum_{i=1}^n w_i t_i = 0$$

$$0 \leq w_i \leq C$$

Решающее правило остается без изменений:

$$\hat{t}(\mathbf{x}) = \text{sign}(\langle \mathbf{z}_*, \mathbf{x} \rangle + b_*) = \text{sign}\left(\sum_{i=1}^n w_i^* t_i \langle \mathbf{x}_i, \mathbf{x} \rangle + b^*\right)$$

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Ядровой переход

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных

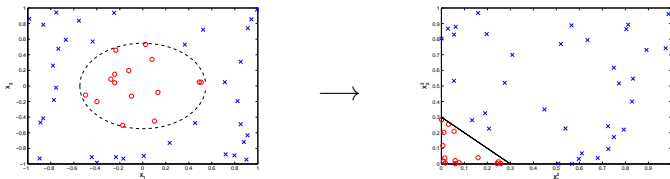
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

- На практике часто встречается ситуация, когда данные порождаются нелинейной разделяющей поверхностью.
- Для обобщения метода на нелинейный случай заметим, что объекты обучающей выборки входят в двойственную задачу оптимизации только в виде попарных скалярных произведений $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$.
- Предположим, что исходное признаковое пространство было подвергнуто некоторому нелинейному преобразованию:

$$\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$$



Ядровой переход

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаваемое распознавание образов

- Для того, чтобы построить гиперплоскость с максимальным зазором в новом пространстве \mathcal{H} необходимо знать лишь $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle_{\mathcal{H}}$.
- Допустим, что существует некоторая «ядровая функция» $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, такая что

$$K(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{H}}$$

- Для построения гиперплоскости с максимальным зазором в пространстве \mathcal{H} нет необходимости задавать преобразование Φ в явном виде, достаточно лишь знать K !
- Задача оптимизации зависит только от попарных скалярных произведений, а решающее правило может быть представлено как

$$\hat{t}(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^n w_i t_i \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}) \rangle_{\mathcal{H}} + b \right) = \text{sign} \left(\sum_{i=1}^n w_i t_i K(\mathbf{x}_i, \mathbf{x}) + b \right)$$

Требования к ядровой функции

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделенных данных
Случай линейно неразделенных данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

Очевидно, что не для любой функции двух переменных K найдутся такие (\mathcal{H}, Φ) , для которых K будет определять скалярное произведение. Необходимыми и достаточными требованиями являются:

- Симметричность

$$K(\mathbf{x}, \mathbf{y}) = K(\mathbf{y}, \mathbf{x})$$

- Неотрицательная определенность (условие Мерсера)

$$\forall g(\mathbf{x}) : \int g^2(\mathbf{x}) d\mathbf{x} < \infty$$

$$\int K(\mathbf{x}, \mathbf{y}) g(\mathbf{x}) g(\mathbf{y}) d\mathbf{x} d\mathbf{y} \geq 0$$

Для фиксированной функции K евклидово пространство \mathcal{H} и преобразование Φ определено не однозначно.

Примеры ядерных функций

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных

Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

- Линейная ядровая функция

$$K(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle + \theta, \theta \geq 0$$

- Полиномиальная ядровая функция

$$K(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + \theta)^d, \theta \geq 0, d \in \mathbb{N}$$

- Гауссиана

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \sigma > 0$$

- Сигмоидная ядровая функция

$$K(\mathbf{x}, \mathbf{y}) = \tanh(\langle \mathbf{x}, \mathbf{y} \rangle + r), r \in \mathbb{R}$$

Это семейство не удовлетворяет условию Мерсера!

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

① Ликбез

② Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

③ Метод опорных векторов для задачи регрессии, [2]

④ Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Пример использования метода опорных векторов

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

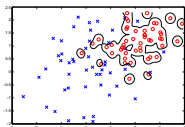
Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

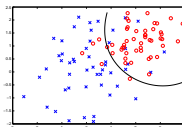
Метод опорных векторов для задачи регрессии, [2]

Вспризнаковое распознавание образов

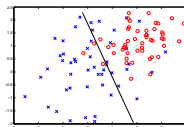
Зависимость от ширины гауссианы



$$C = 1, \sigma^2 = 0.1$$

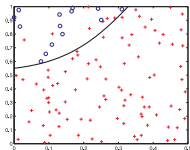


$$C = 1, \sigma^2 = 2$$

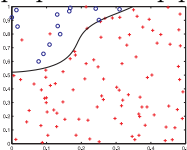


$$C = 1, \sigma^2 = 1000$$

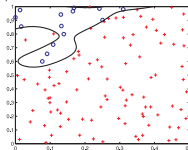
Зависимость от штрафного коэффициента



$$C = 10^{-2}$$



$$C = 1$$



$$C = 10^5$$

Глобальность и единственность решения

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаваемое распознавание образов

- Задача обучения SVM — задача квадратичного программирования
- Известно, что для любой задачи выпуклого программирования (в частности, квадратичного) любой локальный максимум является и глобальным. Кроме того, решение будет единственным, если целевая функция строго вогнута (гессиян отрицательно определен).
- Для обучения SVM можно воспользоваться любым стандартным методом решения задачи квадратичного программирования, однако лучше использовать специальные алгоритмы, учитывающие особенности задачи квадратичного программирования в SVM (например, *SMO* или *SVM^{light}*).

<http://www.kernel-machines.org>

Задача обучения SVM как задача максимума регуляризованного правдоподобия

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаваемое распознавание образов

$$\frac{1}{2} \|\mathbf{z}\|^2 + C \sum_{i=1}^n \xi_i \rightarrow \min_{\mathbf{z}, b, \xi} \quad (5)$$

$$t_i (\langle \mathbf{z}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i \quad (6)$$

$$\xi_i \geq 0 \quad (7)$$

Если $t_i y(\mathbf{x}_i) \geq 1$, то $\xi_i = 0$. Для остальных точек $\xi_i = 1 - t_i y(\mathbf{x}_i)$. Следовательно, функцию (5) можно переписать в виде

$$\sum_{i=1}^n E_{SV}(t_i y(\mathbf{x}_i)) + \lambda \|\mathbf{z}\|^2$$

Здесь $\lambda = (2C)^{-1}$, а $E_{SV}(\cdot)$ — функция потерь, определенная как

$$E_{SV}(s) = [1 - s]_+$$

SVM vs. Логистическая регрессия

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

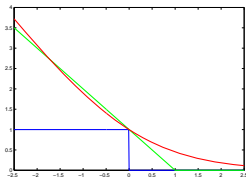
Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов



Логистическая регрессия:

$$\sum_{i=1}^n E_{LR}(t_i y(\mathbf{x}_i)) + \lambda \|\mathbf{w}\|^2 \rightarrow \min_{\mathbf{w}}$$

Здесь $E_{LR}(s) = \log(1 + \exp(-s))$.

SVM:

$$\sum_{i=1}^n E_{SV}(t_i y(\mathbf{x}_i)) + \lambda \|\mathbf{z}\|^2 \rightarrow \min_{\mathbf{z}}$$

Достоинства и недостатки SVM

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций
Случай линейно разделимых данных
Случай линейно неразделимых данных
Ядровой переход
Заключительные замечания

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

- Высокое качество распознавания за счет построения нелинейных разделяющих поверхностей, максимизирующих зазор
- Глобальность и в ряде случаев единственность получаемого решения
- Низкая скорость обучения и большие требования к памяти для задач больших размерностей
- Необходимость грамотного выбора штрафного коэффициента C и параметров ядровой функции

Линейная регрессия vs. SVR

Лекция 3. Метод опорных векторов

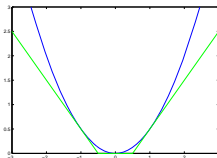
Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов



Линейная регрессия

$$\frac{1}{2} \sum_{i=1}^n (t_i - y(\mathbf{x}_i))^2 + \frac{1}{2} \|\mathbf{w}\|^2 \rightarrow \min_{\mathbf{w}}$$

Для того, чтобы добиться разреженного решения, заменим квадратичную функцию потерь на ε -нечувствительную:

$$E_{\varepsilon}(t - y(\mathbf{x})) = \begin{cases} 0, & \text{если } |t - y(\mathbf{x})| < \varepsilon \\ |t - y(\mathbf{x})| - \varepsilon, & \text{иначе} \end{cases}$$

Тогда мы приходим к следующей оптимизационной задаче:

$$C \sum_{i=1}^n E_{\varepsilon}(y(\mathbf{x}_i) - t_i) + \frac{1}{2} \|\mathbf{z}\|^2 \rightarrow \min_{\mathbf{z}}$$

Ослабляющие коэффициенты

Лекция 3. Метод опорных векторов

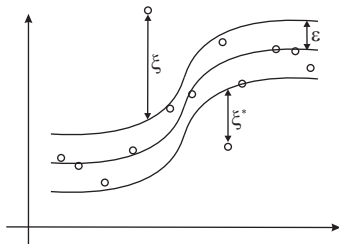
Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов



$$C \sum_{i=1}^n (\xi_i + \xi_i^*) + \frac{1}{2} \|z\|^2 \rightarrow \min_{z, b, \xi, \xi^*}$$

$$t_i \leq y(\mathbf{x}_i) + \varepsilon + \xi_i$$

$$t_i \geq y(\mathbf{x}_i) - \varepsilon - \xi_i^*$$

$$\xi_i, \xi_i^* \geq 0$$

Двойственная задача (Упр.)

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознающее распознавание образов

$$-\frac{1}{2} \sum_{i,j=1}^n (w_i - w_i^*)(w_j - w_j^*)K(\mathbf{x}_i, \mathbf{x}_j) - \varepsilon \sum_{i=1}^n (w_i + w_i^*) + \sum_{i=1}^n t_i (w_i - w_i^*) \rightarrow \max_{\mathbf{w}, \mathbf{w}^*}$$

$$\sum_{i=1}^n (w_i - w_i^*) = 0$$

$$0 \leq w_i, w_i^* \leq C$$

Функция регрессии

$$y(\mathbf{x}) = \sum_{i=1}^n (w_i - w_i^*)K(\mathbf{x}, \mathbf{x}_i) + b$$

Условия дополняющей нежесткости

Лекция 3. Метод опорных векторов

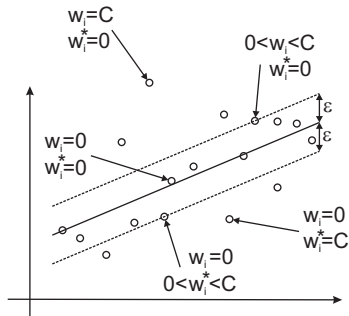
Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов



Условия дополняющей нежесткости

$$w_i(\varepsilon + \xi_i - t_i + \langle \mathbf{z}, \mathbf{x}_i \rangle + b) = 0$$

$$w_i(\varepsilon + \xi_i^* + t_i - \langle \mathbf{z}, \mathbf{x}_i \rangle - b) = 0$$

$$(C - w_i)\xi_i = 0, \quad (C - w_i^*)\xi_i^* = 0$$

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

1 Ликбез

2 Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

3 Метод опорных векторов для задачи регрессии, [2]

4 Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Задачи беспризнакового распознавания образов

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Существует ряд задач распознавания образов, в которых трудно выбрать признаковое пространство, однако, относительно легко ввести меру сходства или несходства между парами объектов. Примеры:

- Задача распознавания личности по фотопортрету
- Задача идентификации личности по подписи в процессе ее формирования
- Задача распознавания классов пространственной структуры белков по последовательностям составляющих их аминокислот

Беспризнаковое распознавание образов: основная идея

Лекция 3. Метод
опорных
векторов

Ветров,
Журавлев

Ликбез

Метод опорных
векторов для
задачи
классификации,
[2]

Метод опорных
векторов для
задачи
регрессии, [2]

Беспризнаковое
распознавание
образов

Основная
методика
беспризнакового
распознавания
образов

Построение
функции,
задающей
скалярное
произведение

- Предположим, что объекты выборки $\omega_1, \dots, \omega_n \in \Omega$
- Пространство Ω является гильбертовым, т.е. на нем определены операции суммы и произведения на число, удовлетворяющие аксиомам линейного пространства:

$$\forall \alpha_1, \alpha_2 \in \Omega \exists \alpha = \alpha_1 + \alpha_2 \in \Omega$$

$$\forall \alpha_1 \in \Omega, c \in \mathbb{R} \exists \alpha = c\alpha_1 \in \Omega$$

1. $\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1$
 2. $\alpha_1 + (\alpha_2 + \alpha_3) = (\alpha_1 + \alpha_2) + \alpha_3$
 3. $\exists \phi \in \Omega : \alpha + \phi = \phi + \alpha = \alpha$
 4. $\forall \alpha \exists (-\alpha) : \alpha + (-\alpha) = \phi$
 5. $c(\alpha_1 + \alpha_2) = c\alpha_1 + c\alpha_2$
 6. $(c + d)\alpha = c\alpha + d\alpha$
 7. $(cd)\alpha = c(d\alpha)$
 8. $1\alpha = \alpha$
- Существует функция $K : \Omega \times \Omega \rightarrow \mathbb{R}$, определяющая скалярное произведение в пространстве Ω :

1. $K(\alpha_1, \alpha_2) = K(\alpha_2, \alpha_1)$
2. $K(\alpha, \alpha) \geq 0, = 0 \Leftrightarrow \alpha = \phi$
3. $K(\alpha_1 + \alpha_2, \alpha) = K(\alpha_1, \alpha) + K(\alpha_2, \alpha)$
4. $K(c\alpha_1, \alpha_2) = cK(\alpha_1, \alpha_2)$

Беспризнаковое распознавание образов: основная идея

Лекция 3. Метод
опорных
векторов

Ветров,
Журавлев

Ликбез

Метод опорных
векторов для
задачи
классификации,
[2]

Метод опорных
векторов для
задачи
регрессии, [2]

Беспризнаковое
распознавание
образов

Основная
методика
беспризнакового
распознавания
образов

Построение
функции,
задающей
скалярное
произведение

Решающая функция

$$\hat{t}(\omega) = \text{sign}(K(\vartheta, \omega) + b)$$

Задача оптимизации

$$\begin{aligned} \frac{1}{2}K(\vartheta, \vartheta) + C \sum_{i=1}^n \xi_i &\rightarrow \min_{\vartheta, b, \xi} \\ t_i(K(\vartheta, \omega_i) + b) &\geq 1 - \xi_i \\ \xi_i &\geq 0 \end{aligned}$$

Двойственная задача
оптимизации

$$\begin{aligned} \sum_{i=1}^n w_i - \frac{1}{2} \sum_{i,j=1}^n t_i t_j w_i w_j K(\omega_i, \omega_j) &\rightarrow \max_w \\ \sum_{i=1}^n t_i w_i &= 0 \\ 0 \leq w_i &\leq C \end{aligned}$$

Решение

$$\hat{t}(\omega) = \text{sign} \left(\sum_{i=1}^n t_i w_i K(\omega_i, \omega) + b \right)$$

Беспризнаковое распознавание образов: основная идея

Лекция 3. Метод
опорных
векторов

Ветров,
Журавлев

Ликбез

Метод опорных
векторов для
задачи
классификации,
[2]

Метод опорных
векторов для
задачи
регрессии, [2]

Беспризнаковое
распознавание
образов

Основная
методика
беспризнакового
распознавания
образов

Построение
функции,
задающей
скалярное
произведение

- Для решения задачи нет необходимости явно задавать линейные операции в пространстве Ω , достаточно лишь потребовать их существование в пространстве, где функция K задает скалярное произведение
- Для применения данного подхода достаточно задать функцию K . Эта функция может быть построена непосредственно, либо с использованием меры сходства

План лекции

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

① Ликбез

② Метод опорных векторов для задачи классификации, [2]

Метод потенциальных функций

Случай линейно разделимых данных

Случай линейно неразделимых данных

Ядровой переход

Заключительные замечания

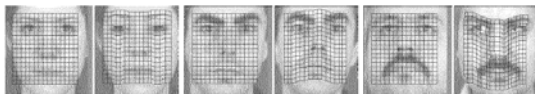
③ Метод опорных векторов для задачи регрессии, [2]

④ Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Явное задание функции K . Пример. [3]



- Рассмотрим задачу распознавания личности по фотопортрету.
- Пусть два изображения заданы векторами яркости в точках раstra: $\mathbf{y}' = \mathbf{y}(\omega') = (y'_t, t \in T)$ и $\mathbf{y}'' = \mathbf{y}(\omega'') = (y''_t, t \in T)$, $T = \{t = (t_1, t_2), t_1 = \overline{1, n_1}, t_2 = \overline{1, n_2}\}$
- Потенциальная функция может быть задана как:
 - $K(\mathbf{y}', \mathbf{y}'') = \langle \mathbf{y}', \mathbf{y}'' \rangle = \sum_{t \in T} y'_t y''_t$
 - $K(\mathbf{y}', \mathbf{y}'') = [\langle \mathbf{y}', \mathbf{y}'' \rangle + 1]^\alpha$
 - $K(\mathbf{y}', \mathbf{y}'') = \exp(-\alpha \|\mathbf{y}' - \mathbf{y}''\|^2) = \exp(-\alpha[\langle \mathbf{y}', \mathbf{y}' \rangle + \langle \mathbf{y}'', \mathbf{y}'' \rangle - 2\langle \mathbf{y}', \mathbf{y}'' \rangle])$
- Пусть задана эластичная деформация раstra $\mathbf{t} \rightarrow \mathbf{t} + \mathbf{x}_t$.
Эластичная потенциальная функция:

$$K(\mathbf{y}', \mathbf{y}'') = \sum_{t \in T} y'_t y''_{t+\mathbf{x}_t}$$

Построение функции K с использованием метрики

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознающее распознавание образов

Основная методика беспознающего распознавания образов

Построение функции, задающей скалярное произведение

- Пусть задано метрическое пространство A с метрикой $\rho : A \times A \rightarrow \mathbb{R}$:

- $\rho(\alpha_1, \alpha_2) \geq 0, = 0 \Leftrightarrow \alpha_1 = \alpha_2$
- $\rho(\alpha_1, \alpha_2) = \rho(\alpha_2, \alpha_1)$
- $\rho(\alpha_1, \alpha_3) \leq \rho(\alpha_1, \alpha_2) + \rho(\alpha_2, \alpha_3)$

- Общностью двух элементов из A относительно некоторого центра $\phi \in A$ назовем следующую величину:

$$\mu_\phi(\alpha_1, \alpha_2) = \frac{1}{2} \left[\rho^2(\alpha_1, \phi) + \rho^2(\alpha_2, \phi) - \rho^2(\alpha_1, \alpha_2) \right]$$

- Свойства общности

- $\mu_\phi(\alpha_1, \alpha_2) = \mu_\phi(\alpha_2, \alpha_1)$
- $\mu_\phi(\alpha, \alpha) \geq 0 = 0 \Leftrightarrow \alpha = \phi$
- $\forall \alpha \in A \mu_\phi(\alpha, \phi) = 0$
- $\mu_\phi(\alpha, \alpha) = \rho^2(\alpha, \phi)$
- $\rho(\alpha_1, \alpha_2) = (\mu_\phi(\alpha_1, \alpha_1) + \mu_\phi(\alpha_2, \alpha_2) - 2\mu_\phi(\alpha_1, \alpha_2))^{1/2}$
- $\mu_\phi(\alpha_1, \alpha_2) \leq \frac{1}{2} [\mu_\phi(\alpha_1, \alpha_1) + \mu_\phi(\alpha_2, \alpha_2)]$
- $|\mu_\phi(\alpha_1, \alpha_2)| \leq \sqrt{\mu_\phi(\alpha_1, \alpha_1)} \sqrt{\mu_\phi(\alpha_2, \alpha_2)}$
- $\mu_\phi(\alpha_1, \alpha_2) = \mu_\phi(\alpha_1, \alpha_2) - \mu_\phi(\alpha_1, \tilde{\phi}) - \mu_\phi(\alpha_2, \tilde{\phi}) + \mu_\phi(\tilde{\phi}, \tilde{\phi})$

- По своим свойствам общность очень похожа на скалярное произведение!

Матрица общностей

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

- Пусть $\{\alpha_1, \dots, \alpha_q\} \subset A$ — конечная совокупность элементов метрического пространства с центром $\phi \in A$. Составим матрицу общностей этих элементов:

$$M_\phi = (\mu_\phi(\alpha_i, \alpha_j), i, j = 1, \dots, q)$$

- Матрица общностей является симметричной, диагональные элементы неотрицательны.
- В отличие от матрицы скалярных произведений, которая всегда неотрицательно определена, матрица общностей может иметь собственные значения любого знака

Теорема

Если M_ϕ является неотрицательно определенной для любой конечной совокупности элементов, то матрица $M_{\tilde{\phi}}$ относительно другого центра $\tilde{\phi}$ также является неотрицательно определенной.

Соосность элементов

Лекция 3. Метод опорных векторов

Ветров, Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

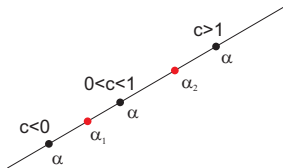
Беспризнаковое распознавание образов

Основная методика непризнакового распознавания образов

Построение функции, задающей скалярное произведение

Элемент $\alpha \in A$ называется **соосным элементом** для упорядоченной пары $[\alpha_1, \alpha_2]$ с коэффициентом $c \in \mathbb{R}$ и обозначается $\alpha = \text{соах}([\alpha_1, \alpha_2]; c)$, если

$$\rho(\alpha_1, \alpha) = |c|\rho(\alpha_1, \alpha_2), \quad \rho(\alpha_2, \alpha) = |c-1|\rho(\alpha_1, \alpha_2)$$



Очевидно, что $\alpha_1 = \text{соах}([\alpha_1, \alpha_2]; 0)$, $\alpha_2 = \text{соах}([\alpha_1, \alpha_2]; 1)$. Если $\alpha = \text{соах}([\alpha_1, \alpha_2]; c)$, то $\alpha = \text{соах}([\alpha_2, \alpha_1]; 1 - c)$. Для элементов $[\alpha_1, \alpha_2]$ и $\alpha = \text{соах}([\alpha_1, \alpha_2]; c)$ неравенство треугольника переходит в равенство:

$$\rho(\alpha_1, \alpha_2) + \rho(\alpha_2, \alpha) = \rho(\alpha_1, \alpha), \quad \text{если } c > 1$$

$$\rho(\alpha_1, \alpha) + \rho(\alpha, \alpha_2) = \rho(\alpha_1, \alpha_2), \quad \text{если } 0 < c \leq 1$$

$$\rho(\alpha, \alpha_1) + \rho(\alpha_1, \alpha_2) = \rho(\alpha, \alpha_2), \quad \text{если } c \leq 0$$

Введение линейных операций

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспознаковое распознавание образов

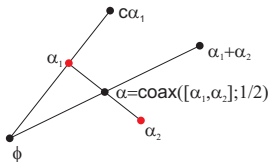
Основная методика беспознакового распознавания образов

Построение функции, задающей скалярное произведение

Метрическое пространство A называется **евклидовым метрическим пространством**, если $\forall [\alpha_1, \alpha_2], \alpha_1, \alpha_2 \in A$ и $\forall c \in \mathbb{R} \exists \alpha \in A : \alpha = \text{соак}([\alpha_1, \alpha_2]; c)$ и матрица общности всякой конечной совокупности элементов из A является неотрицательно определенной.

Введем операции суммы и умножения на число:

$$c\alpha \triangleq \text{соак}([\phi, \alpha]; c) \quad \alpha_1 + \alpha_2 = 2 \text{соак}([\alpha_1, \alpha_2]; \frac{1}{2})$$



Свойства введенных операций

Лекция 3. Метод опорных векторов

Ветров,
Журавлев

Ликбез

Метод опорных векторов для задачи классификации, [2]

Метод опорных векторов для задачи регрессии, [2]

Беспризнаковое распознавание образов

Основная методика беспризнакового распознавания образов

Построение функции, задающей скалярное произведение

Теорема

В евклидовом метрическом пространстве введенные операции сложения и умножения на число, а также скалярное произведение понимаемое как общность элементов $\langle \alpha_1, \alpha_2 \rangle = \mu_\phi(\alpha_1, \alpha_2)$ удовлетворяют всем требованиям гильбертова пространства:

- $\alpha_1 + \alpha_2 = \alpha_2 + \alpha_1$, $(\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3)$
- $\alpha + \phi = \alpha$, $c\phi = \phi$
- $\forall \alpha \exists (-\alpha) : (-\alpha) + \alpha = \phi$
- $c_1(c_2\alpha) = (c_1c_2)\alpha$
- $1\alpha = \alpha$
- $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$, $c(\alpha_1 + \alpha_2) = c\alpha_1 + c\alpha_2$
- $\langle \alpha_1, \alpha_2 \rangle = \langle \alpha_2, \alpha_1 \rangle$, $\langle \alpha_1 + \alpha_2, \alpha_3 \rangle = \langle \alpha_1, \alpha_3 \rangle + \langle \alpha_2, \alpha_3 \rangle$,
 $\langle c\alpha_1, \alpha_2 \rangle = c\langle \alpha_1, \alpha_2 \rangle$
- $\langle \alpha, \alpha \rangle \geq 0$, $= 0 \Leftrightarrow \alpha = \phi$
- $\|\alpha_1 - \alpha_2\| = \sqrt{\langle \alpha_1 - \alpha_2, \alpha_1 - \alpha_2 \rangle} = \rho(\alpha_1, \alpha_2)$

Полезная литература

Лекция 3. Метод
опорных
векторов

Ветров,
Журавлев

Приложение
Полезная
литература



М. А. Айзерман, Э. М. Браверман, Л. И. Розоноэр
Метод потенциальных функций в теории обучения
машин
М.: Наука, 1970



V. N. Vapnik
The Nature of Statistical Learning Theory
Springer, 1995



О. С. Середин
Методы и алгоритмы беспризнакового распознавания
образов
Дисс. к.ф.-м.н., Тульский гос. университет, 2001