Лекция 3

Метрические методы

3.1. Метод ближайших соседей

3.1. Метод ближайших соседей

Формализация понятия “близости”:

задана ф-ия расстояния *p: X×X→[0,∞)*

Пример: Евклидово расстояние и его обобщение

$$p(x,x_i) = \sqrt{\left(\sum_{j=1}^{n} |x^j - x_i^j|^2\right)} \qquad p(x,x_i) = \sqrt[p]{w_j\left(\sum_{j=1}^{n} |x^j - x_i^j|^p\right)}$$

$x = (x^1, \dots, x^n)$ - вектор признаков объекта x

$x_i = (x_i^1, \dots, x_i^n)$ - вектор признаков объекта x_i

3.1. Метод ближайших соседей

Для произвольного $x \in X$ отранжируем объекты $x_1 \ldots x_l$:

$$\rho(x, x^{(1)}) \leqslant \rho(x, x^{(2)}) \leqslant \cdots \leqslant \rho(x, x^{(\ell)}),$$

$x^{(i)}$ - *i*-й сосед объекта x

$y^{(i)}$ - ответ на *i*-м соседе объекта x

Метрический алгоритм классификации: $a(x; X^\ell) = \arg\max_{y \in Y} \underbrace{\sum_{i=1}^{\ell} [y^{(i)} = y] \, w(i, x)}_{\Gamma_y(x)},$

w(i,x) - вес, оценка сходства x с уго *i*-м соседом, неотрицательная, невозрастающая по *i*

$\Gamma_y(x)$ - оценка близости объекта *x* к классу *y*

3.1. Метод ближайших соседей

$w(i,x) = [i \le k]$.

$w(i,x) = [i \le 1]$ - метод ближайшего соседа.

Преимущества:

- простота реализации
- параметр k можно оптимизировать по критерию скользящего контроля (leave-one-out)

$$\mathrm{LOO}(k, X^\ell) = \sum_{i=1}^{\ell} \Big[a(x_i; X^\ell \backslash \{x_i\}, k) \ne y_i \Big] \to \min_k$$

Проблемы:

- возможны ситуации, когда классификация неоднозначна: $\Gamma_y(x) = \Gamma_s(x)$ для пары классов $y \ne s$
- учитываются не значения расстояний, а только их ранги

3.1. Метод окна Парзена

3.2. Метод окна Парзена

$w(i,x) = K\left(\frac{\rho(x,x^{(i)})}{h}\right)$, где h - ширина окна

K(r) - ядро, не возрастает и положительно на [0, 1]

Метод Парзоновского окна фиксированной ширины:

$$a(x; X^\ell, h, K) = \arg\max_{y \in Y} \sum_{i=1}^{\ell} [y_i = y]\, K\left(\frac{\rho(x, x_i)}{h}\right)$$

Метод Парзоновского окна переменной ширины:

$$a(x; X^\ell, k, K) = \arg\max_{y \in Y} \sum_{i=1}^{\ell} [y_i = y]\, K\left(\frac{\rho(x, x_i)}{\rho(x, x^{(k+1)})}\right)$$

Оптимизация параметров по критерию leave-one-out

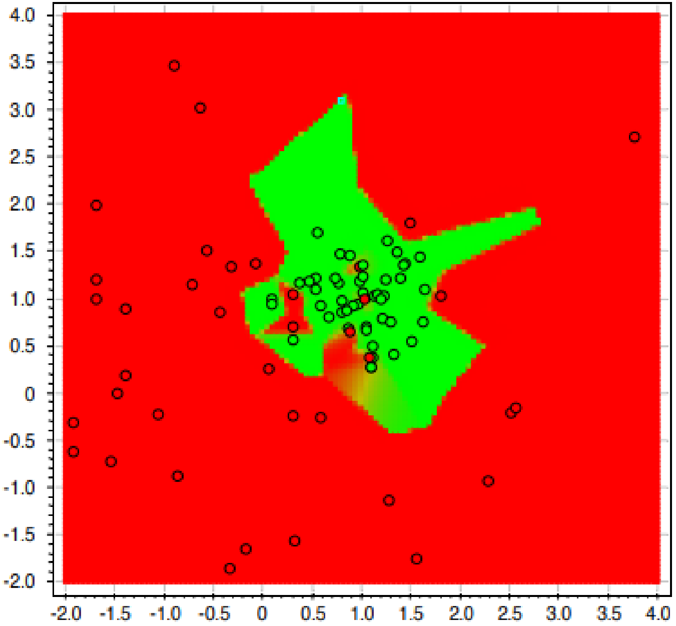
- выбор ширины окна h или числа соседей k;
- выбор ядра K слабо влияет на качество классификации

3.2. Метод окна Парзена

Пример: двумерная выборка. два класса Y={-1,+1}

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x) = \operatorname{sign}\big(\underbrace{\Gamma_{+1}(x) - \Gamma_{-1}(x)}\big)$$

$h = 0.05$

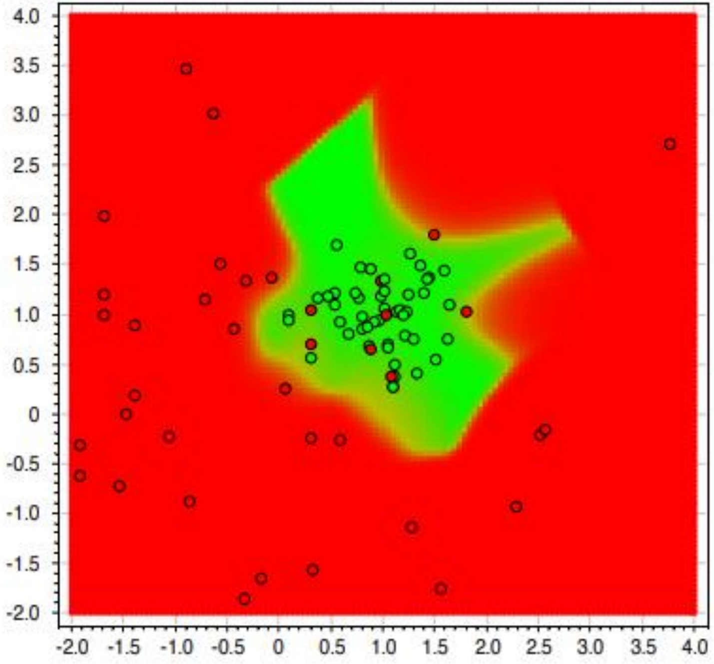


3.2. Метод окна Парзена

Пример: двумерная выборка. два класса Y={-1,+1}

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x) = \operatorname{sign}\big(\underbrace{\Gamma_{+1}(x) - \Gamma_{-1}(x)}\big)$$

$h = 0.2$

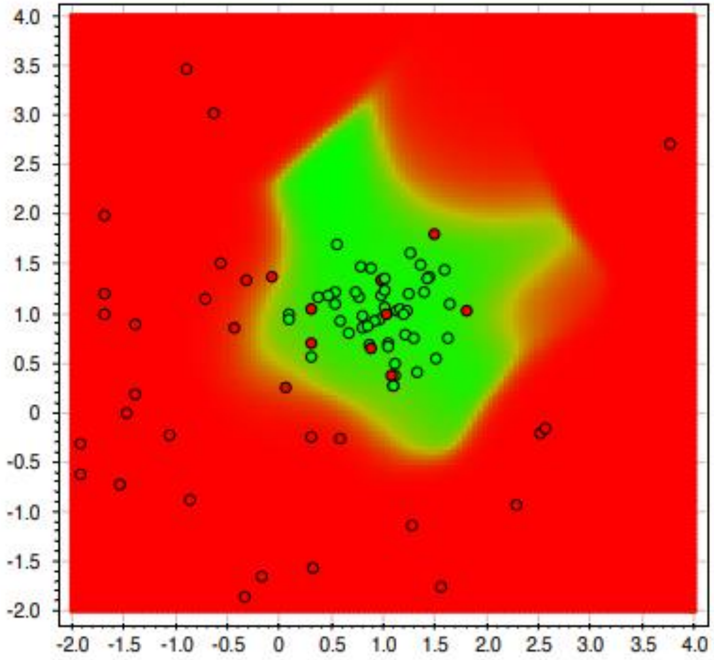


3.2. Метод окна Парзена

Пример: двумерная выборка. два класса Y={-1,+1}

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x) = \mathrm{sign}\big(\underbrace{\Gamma_{+1}(x) - \Gamma_{-1}(x)}\big)$$

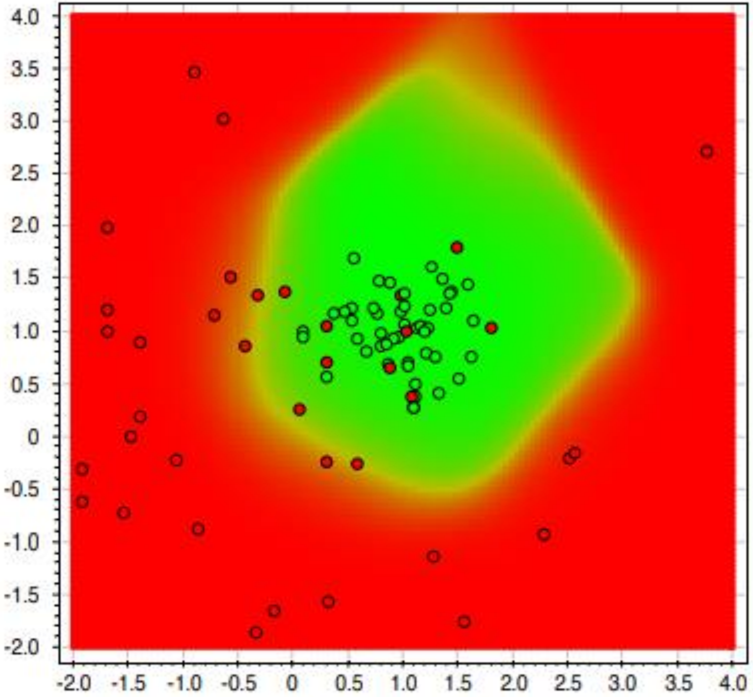
$h = 0.3$



3.2. Метод окна Парзена

Пример: двумерная выборка. два класса Y={-1,+1}

$h = 0.5$

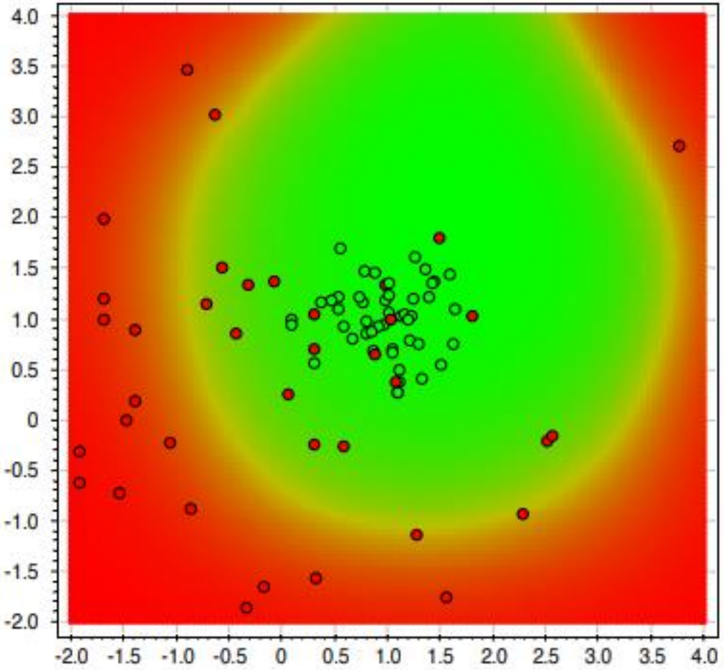


3.2. Метод окна Парзена

Пример: двумерная выборка. два класса Y={-1,+1}

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x) = \operatorname{sign}\underbrace{\left(\Gamma_{+1}(x) - \Gamma_{-1}(x)\right)}$$

$h = 1.0$

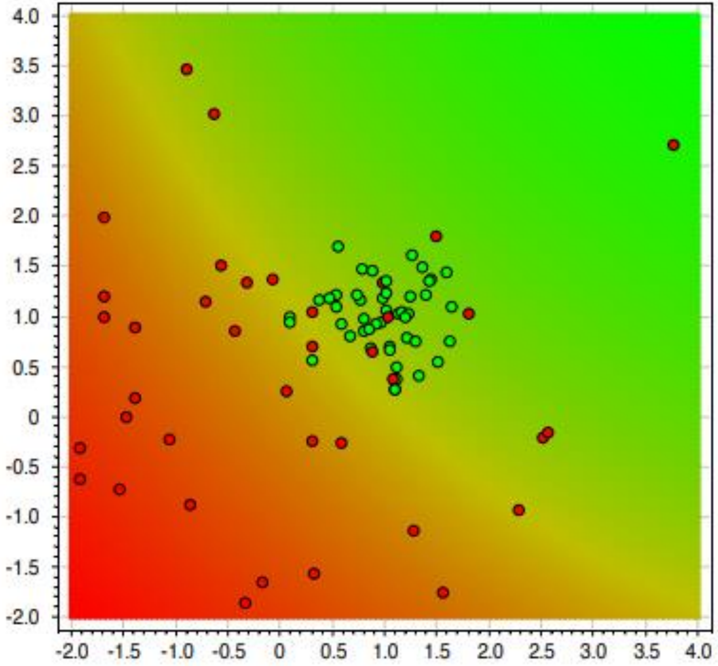


3.2. Метод окна Парзена

Пример: двумерная выборка, два класса Y={-1,+1}

$$a(x) = \arg\max_{y \in Y} \Gamma_y(x) = \mathrm{sign}\big(\underbrace{\Gamma_{+1}(x) - \Gamma_{-1}(x)}\big)$$

$h = 5.0$



3.2. Метод потенциальных функций

$$w(i,x) = \gamma^{(i)} K\left(\frac{\rho(x,x^{(i)})}{h^{(i)}}\right)$$

Без ранжирования:

$$a(x;X^\ell) = \arg\max_{y\in Y} \sum_{i=1}^{\ell} [y_i = y]\, \gamma_i\, K\left(\frac{\rho(x,x_i)}{h_i}\right)$$

где γ_i - веса объектов, $\gamma_i \geq 0$, $h_i > 0$

Физическая аналогия:

γ_i - величина “заряда” в точке x_i

h_i - радиус действия потенциала в точке x_i

y_i - знак заряда (бинарная классификация Y={-1,+1})

в электростатике $K(r) = \frac{1}{r}$ или $\frac{1}{r+a}$

Для задачи классификации нет таких ограничений на K

3.2. Метод потенциальных функций

Два класса Y={-1.+1}

$$a(x; X^\ell) = \arg\max_{y \in Y} \Gamma_y(x) = \mathrm{sign}\big(\Gamma_{+1}(x) - \Gamma_{-1}(x)\big) =$$

$$= \mathrm{sign} \sum_{i=1}^{\ell} \gamma_i y_i \, K\left(\frac{\rho(x, x_i)}{h_i}\right).$$

Сравним с линейной моделью классификации

$$a(x) = \mathrm{sign} \sum_{j=1}^{n} \gamma_j f_j(x)$$

- функции $f_j(x) = y_j K\left(\frac{1}{h_j} \rho(x, x_j)\right)$ - признаки объекта x
- γ_i - веса линейного классификатора
- n = ℓ - число признаков равно числу объектов обучения

Метрические методы

- Метрически классификаторы - одни из самых простых
  Качество классификации определяется качеством метрики
- Что можно обучать:
  - число ближайших соседей k или величину окна h;
  - веса объектов;
  - набор эталонов;
  - метрику (distance learning, similarity learning);
  - веса признаков;
  - функцию ядра K(r).