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 парзеновского 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 парзеновского окна *переменной ширины:*

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

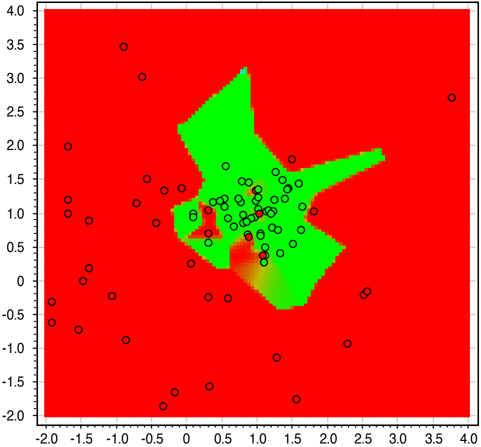
Оптимизация параметров — по критерию LOO:

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

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

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

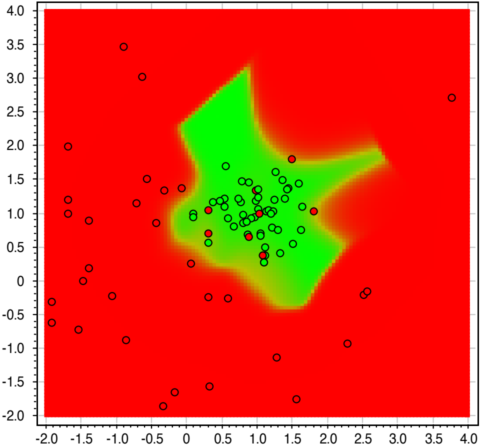
$h = 0.05$



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

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

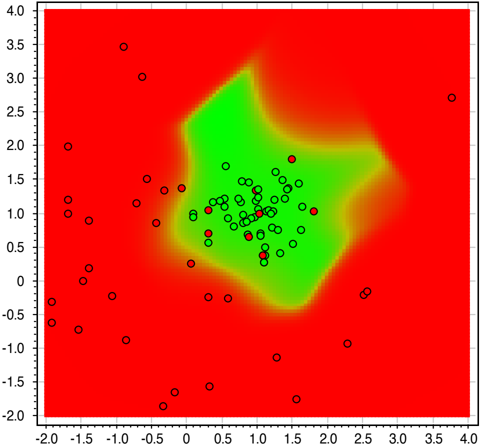
$h = 0.2$



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

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

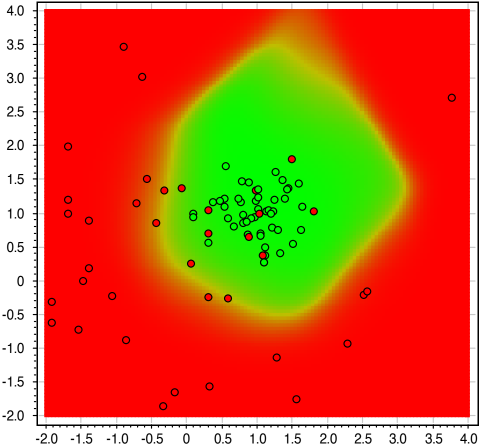
$h = 0.3$



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

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

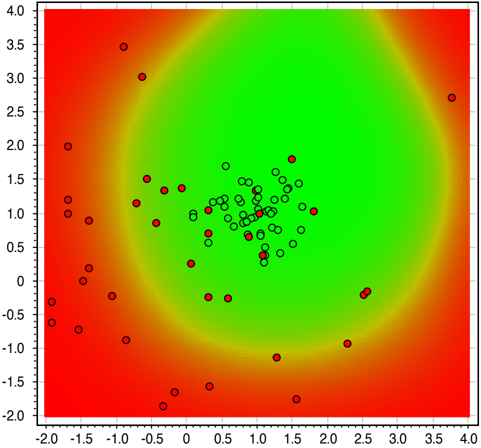
$h = 0.5$



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

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

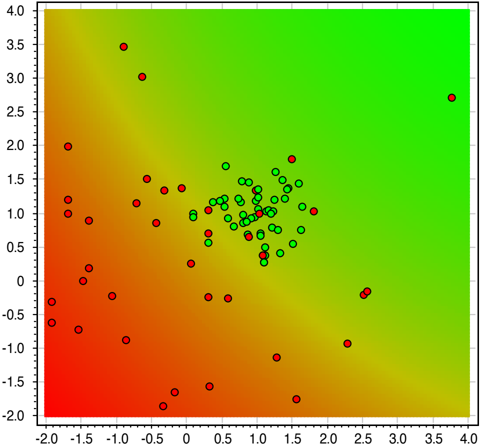
$h = 1.0$



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

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

$h = 5.0$



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

Более простая запись (без ранжирования объектов):

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

где γ_i — веса объектов, $\gamma_i \geqslant 0$, $h_i > 0$.

Физическая аналогия:

γ_i — величина «заряда» в точке x_i;
h_i — «радиус действия» потенциала с центром в точке x_i;
y_i — знак «заряда» (в случае двух классов $Y = \{-1, +1\}$);
в электростатике $K(r) = \frac{1}{r}$ или $\frac{1}{r+a}$,
для задач классификации нет таких ограничений на K.

Два класса: $Y = \{-1, +1\}$.

$$a(x; X^\ell) = \arg\max_{y\in Y} \Gamma_y(x) = \mathrm{sign}\bigl(\Gamma_{+1}(x) - \Gamma_{-1}(x)\bigr) =$$
$$= \mathrm{sign} \sum_{i=1}^{\ell} \gamma_i y_i \, K\left(\frac{\rho(x, x_i)}{h_i}\right).$$

Сравним с линейной моделью классификации:

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

- функции $f_j(x) = y_j K\bigl(\frac{1}{h_j}\rho(x, x_j)\bigr)$ — признаки объекта x
- γ_j — веса линейного классификатора
- $n = \ell$ — число признаков равно числу объектов обучения

Резюме

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