

Parzen window

Парзеновское окно - случай алгоритма kNN взвешенных соседей, где вес определяется формулой:

$$w(i, u) = K \left(\frac{1}{h} \rho(u, x_u^{(i)}) \right)$$

, где ρ - расстояние, $K(z)$ - функция ядра, невозрастающая на $[0; \infty)$. Сам алгоритм имеет вид:

$$a(u; X^l; h) = \underset{y \in Y}{\operatorname{argmax}} \sum_{i: y^{(i)} = y} K \left(\frac{\rho(u, x_u^{(i)})}{h} \right)$$

Оптимальное значение h находим по методу leave-one-out:

$$h = \underset{h}{\operatorname{argmax}} \sum_{i=1}^l \log p_h(x_i; X^m/x_i)$$

Таблица функций ядер, которые были использованы в данном примере для Ирисов Фишера:

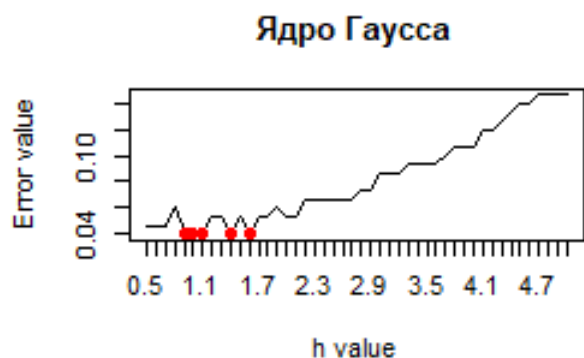
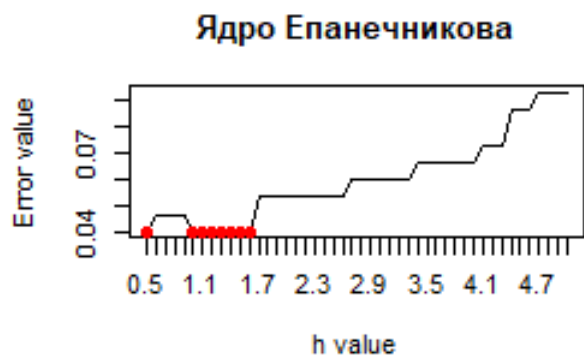
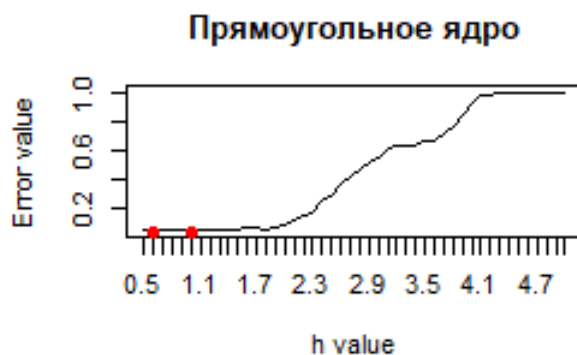
№	Ядро $K(u)$	формула
1	Епанечникова	$K(u) = \frac{3}{4}(1 - u^2)[u \leq 1]$
2	Квартическое	$K(u) = \frac{15}{16}(1 - u^2)^2[u \leq 1]$
3	Треугольное	$K(u) = (1 - u)[u \leq 1]$
4	Гауссовское	$K(u) = (2\pi)^{(\frac{1}{2})} \exp(-\frac{1}{2}u^2)[u \leq 1]$
5	Прямоугольное	$K(u) = \frac{1}{2}[u \leq 1]$

Классификация ведется по 3 и 4 признакам ириса в датафрейме, т.е. по Petal.Length и Petal.Width.

Запишем результаты в таблицу:

№	Ядро	Минимальное кол-во ошибок при LOO	Оптимальный выбор h
1	Епанечникова	6	0.5 и $[1, 1.6]$ с шагом 0.1
2	Квартическое	6	$[0.5, 0.6]$ и $[1.2, 2]$
3	Треугольное	6	$[0.5, 0.6]$ и $[1.1, 2.7]$
4	Гауссовское	6	$[0.9, 1.1]$ и 1.4, 1.6
5	Прямоугольное	6	0.6 и 1

Представим графики зависимости h от кол-ва ошибок (красные точки - минимальное значение ошибки):



Для более точного нахождения h нужно брать шаг меньше, чем 0.1, хотя даже на этих графиках видно, начиная с какого шага точность резко падает.

Итоговая точность для всех ядер с наилучшим h из интервала $[0.5, 5]$, шагом в 0.1 - 96%. Повысить точность можно, используя все 4 признака Ирисов.