

- X — объекты (часто \mathbb{R}^n); Y — ответы (часто \mathbb{R} , реже \mathbb{R}^m);
 $X^\ell = (x_i, y_i)_{i=1}^\ell$ — обучающая выборка;
 $y_i = y(x_i)$, $y: X \rightarrow Y$ — неизвестная зависимость;
- $a(x) = f(x, \alpha)$ — параметрическая модель зависимости,
 $\alpha \in \mathbb{R}^p$ — вектор параметров модели.
- Метод наименьших квадратов (МНК):

$$Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} w_i (f(x_i, \alpha) - y_i)^2 \rightarrow \min_{\alpha},$$

где w_i — вес, степень важности i -го объекта.

- **Недостаток:**
надо иметь хорошую параметрическую модель $f(x, \alpha)$

Приближение константой $f(x, \alpha) = \alpha$ в окрестности $x \in X$:

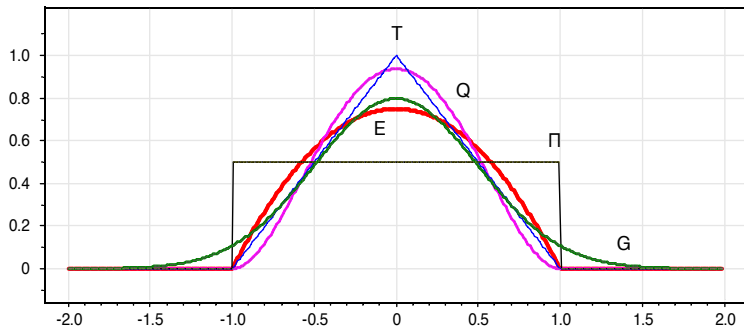
$$Q(\alpha; X^\ell) = \sum_{i=1}^{\ell} w_i(x) (\alpha - y_i)^2 \rightarrow \min_{\alpha \in \mathbb{R}};$$

где $w_i(x) = K\left(\frac{\rho(x, x_i)}{h}\right)$ — веса объектов x_i относительно x ;
 $K(r)$ — ядро, невозрастающее, ограниченное, гладкое;
 h — ширина окна сглаживания.

Формула ядерного сглаживания Надарая–Ватсона:

$$a_h(x; X^\ell) = \frac{\sum_{i=1}^{\ell} y_i w_i(x)}{\sum_{i=1}^{\ell} w_i(x)} = \frac{\sum_{i=1}^{\ell} y_i K\left(\frac{\rho(x, x_i)}{h}\right)}{\sum_{i=1}^{\ell} K\left(\frac{\rho(x, x_i)}{h}\right)}.$$

Часто используемые ядра $K(r)$



$\Pi(r) = [|r| \leq 1]$ — прямоугольное

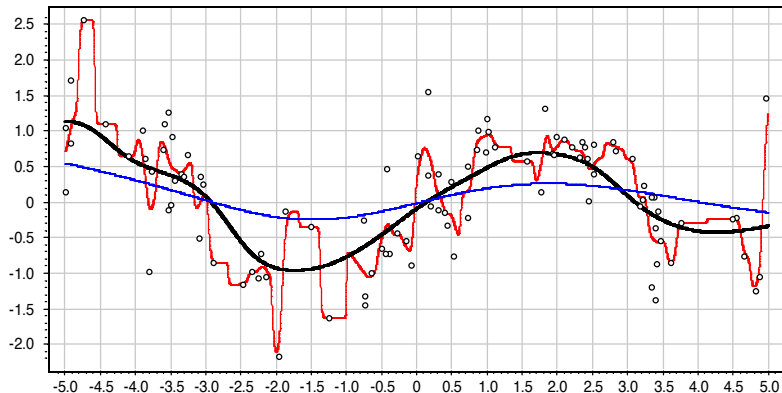
$T(r) = (1 - |r|) [|r| \leq 1]$ — треугольное

$E(r) = (1 - r^2) [|r| \leq 1]$ — квадратичное (Епанечникова)

$Q(r) = (1 - r^2)^2 [|r| \leq 1]$ — четвертое

$G(r) = \exp(-2r^2)$ — гауссовское

$h \in \{0.1, 1.0, 3.0\}$, гауссовское ядро $K(r) = \exp(-2r^2)$

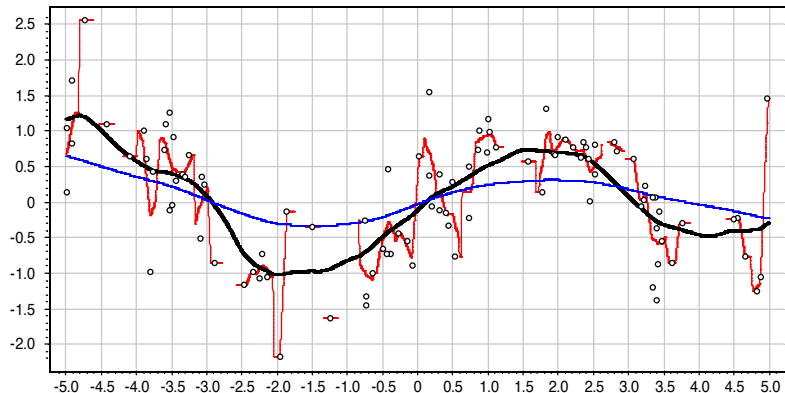


Гауссовское ядро \Rightarrow гладкая аппроксимация

Ширина окна существенно влияет на точность аппроксимации

Выбор ядра K и ширины окна h

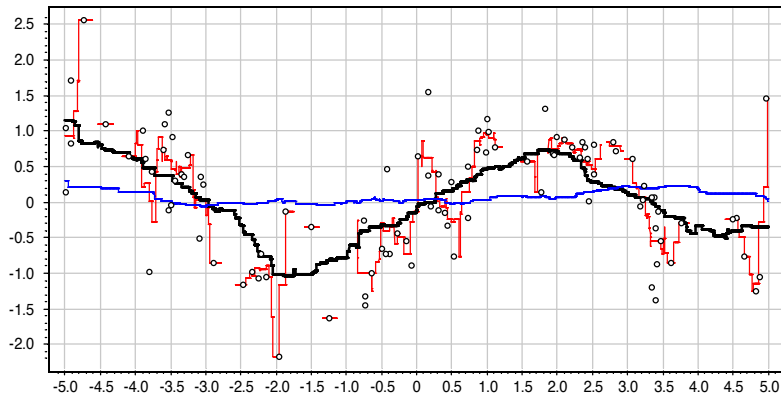
$h \in \{0.1, 1.0, 3.0\}$, треугольное ядро $K(r) = (1 - |r|) [|r| \leq 1]$



Треугольное ядро \Rightarrow кусочно-линейная аппроксимация

Аппроксимация не определена, если в окне нет точек выборки

$h \in \{0.1, 1.0, 3.0\}$, прямоугольное ядро $K(r) = [|r| \leq 1]$



Прямоугольное ядро \Rightarrow кусочно-постоянная аппроксимация
Выбор ядра слабо влияет на точность аппроксимации

- Ядро $K(r)$
 - существенно влияет на гладкость функции $a_h(x)$,
 - слабо влияет на качество аппроксимации.
- Ширина окна h
 - существенно влияет на качество аппроксимации.
- Переменная ширина окна $h(x)$ по k ближайшим соседям:

$$w_i(x) = K\left(\frac{\rho(x, x_i)}{h(x)}\right),$$

где $h(x) = \rho(x, x^{(k+1)})$, $x^{(k)}$ — k -й сосед объекта x .

- Оптимизация ширины окна по скользящему контролю:

$$\text{LOO}(h, X^\ell) = \sum_{i=1}^{\ell} \left(a_h(x_i; X^\ell \setminus \{x_i\}) - y_i \right)^2 \rightarrow \min_h.$$

- Непараметрическая регрессия избегает использования параметрической модели зависимости $f(x, \alpha)$.
- Неявно моделью является функция расстояния $\rho(x, x_i)$ между объектами.
- Что можно обучать:
 - число ближайших соседей k или ширину окна h ;
 - веса объектов (обнаруживать выбросы);
 - метрику (distance learning, similarity learning);
 - в частности, веса признаков в метрике.
- Непараметрическая регрессия часто используется как инструмент предварительной обработки данных для сглаживания шумов в данных.