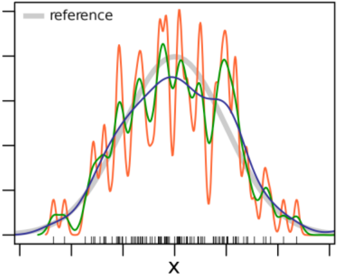
Ядерная оценка плотности

Виктор Китов

victorkitov.github.io



Курс поддержан фондом 'Интеллект'

Победитель конкурса VK среди курсов по IT

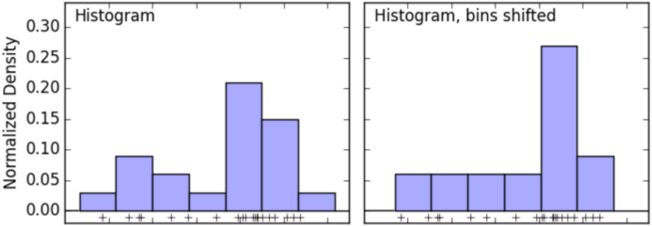
Содержание

Непараметрическая оценка плотности

- Нужна непараметрическая оценка плотности.

Непараметрическая оценка плотности

- Нужна непараметрическая оценка плотности.
- Гистограмма - решение, но она
  - кусочно-постоянная
  - выбор диапазона ячеек влияет на результат:



Идея ядерной оценки плотности

$$p(x) = \lim_{h \to 0} \frac{1}{2h} P\left(\xi \in [x-h, x+h]\right)$$

Идея ядерной оценки плотности

$$p(x) = \lim_{h\to 0}\frac{1}{2h}P\left(\xi \in [x-h, x+h]\right)$$

Частотная оценка:

$$\widehat{p}\left(x\right) = \frac{1}{2h}\frac{1}{N}\sum_{n=1}^{N}\mathbb{I}\left\{x_n \in [x-h, x+h]\right\}$$

Идея ядерной оценки плотности

$$p(x) = \lim_{h\to 0} \frac{1}{2h} P\left(\xi \in [x-h, x+h]\right)$$

Частотная оценка:

$$\widehat{p}\,(x) = \frac{1}{2h}\frac{1}{N}\sum_{n=1}^{N}\mathbb{I}\left\{x_n \in [x-h, x+h]\right\}$$

$$= \frac{1}{Nh}\sum_{n=1}^{N}\frac{1}{2}\mathbb{I}\left[|x-x_n| \le h\right] = \frac{1}{Nh}\sum_{n=1}^{N}\frac{1}{2}\mathbb{I}\left[\frac{|x-x_n|}{h} \le 1\right]$$

Идея ядерной оценки плотности

$$p(x) = \lim_{h \to 0} \frac{1}{2h} P\left(\xi \in [x-h, x+h]\right)$$

Частотная оценка:

$$\widehat{p}\left(x\right) = \frac{1}{2h}\frac{1}{N}\sum_{n=1}^{N}\mathbb{I}\left\{x_n \in [x-h, x+h]\right\}$$

$$= \frac{1}{Nh}\sum_{n=1}^{N}\frac{1}{2}\mathbb{I}\left[|x-x_n| \le h\right] = \frac{1}{Nh}\sum_{n=1}^{N}\frac{1}{2}\mathbb{I}\left[\frac{|x-x_n|}{h} \le 1\right]$$

$$= \frac{1}{Nh}\sum_{n=1}^{N}K\left(\frac{x-x_n}{h}\right), \quad K\left(u\right) = \frac{1}{2}\mathbb{I}[|u| \le 1]$$

Ядерная оценка плотности

Ядерная оценка плотности

Англ. Kernel density estimation (KDE)

$$\widehat{p}(x) = \frac{1}{Nh}\sum_{n=1}^{N} K\left(\frac{x - x_n}{h}\right)$$

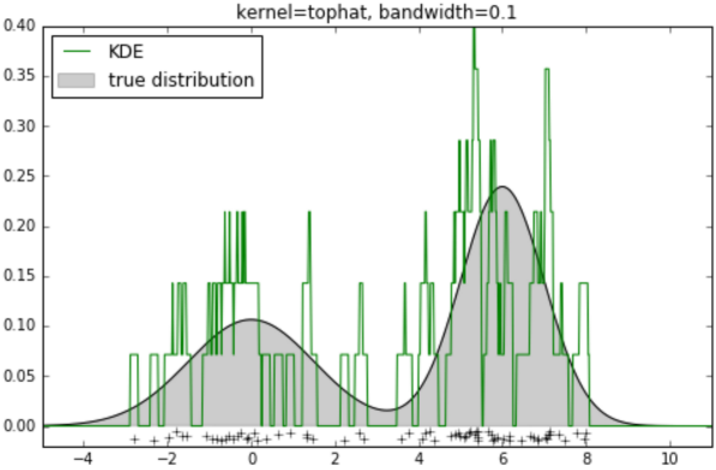
Для ф-ции ядра (kernel), удовлетворяющей

$$K(u) \geq 0, \quad \int_{-\infty}^{+\infty} K(u)du = 1$$

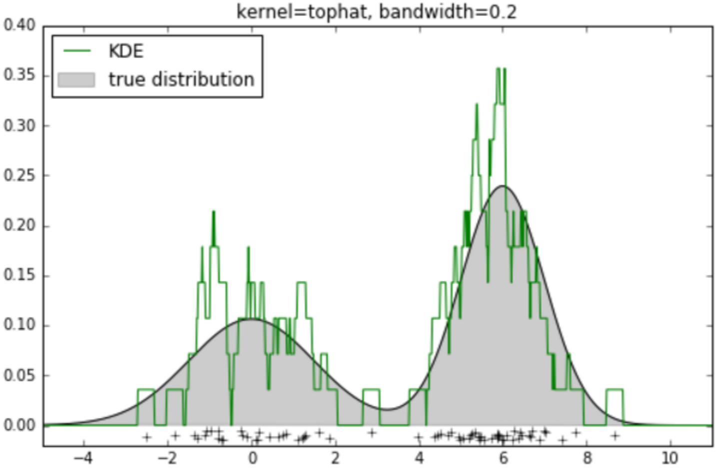
- пример: $K(u) = \frac{1}{2}\mathbb{I}[|u| \leq 1]$ - *прямоугольное (tophat)* ядро
- h - параметр ширины окна *(bandwidth)*
- h контролирует гладкость[1]

[1] каким образом?

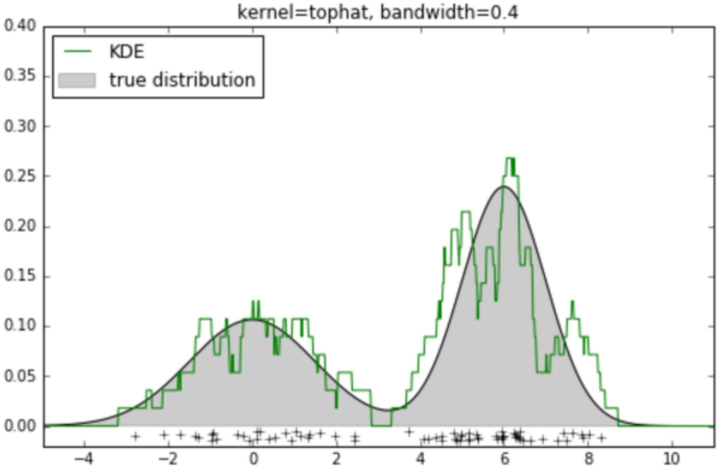
Пример: ядерная оценка с tophat ядром



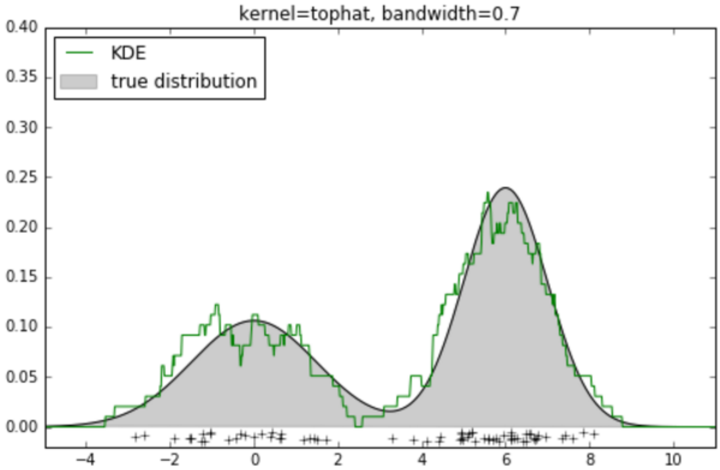
Пример: ядерная оценка с tophat ядром



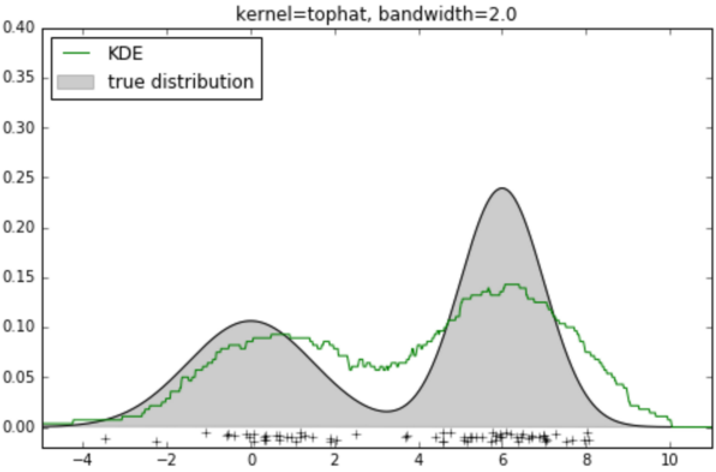
Пример: ядерная оценка с tophat ядром



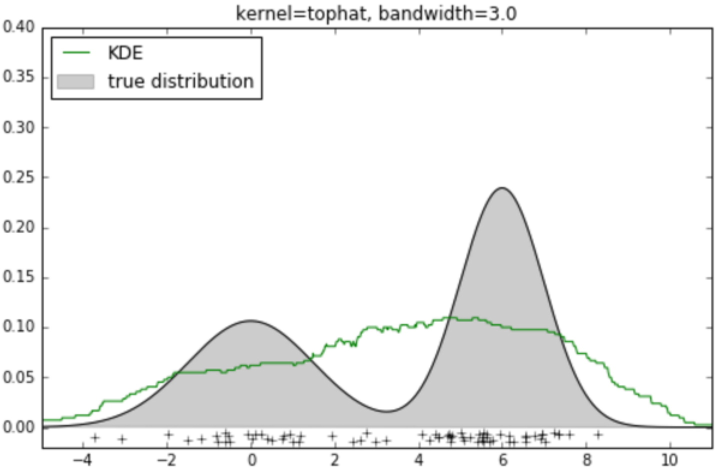
Пример: ядерная оценка с tophat ядром



Пример: ядерная оценка с tophat ядром



Пример: ядерная оценка с tophat ядром



Другие функции ядра

Проблемы top-hat ядра:

- Результирующая оценка $\widehat{p}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right)$ кусочно-постоянна.
- Влияние точки x_i не изменяется с $\rho(x, x_i)$ вблизи x.

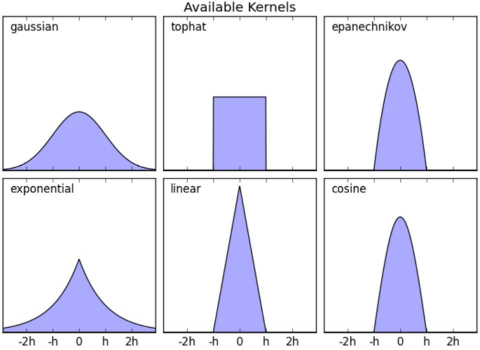
[2] Picture source.

Другие функции ядра

Проблемы top-hat ядра:

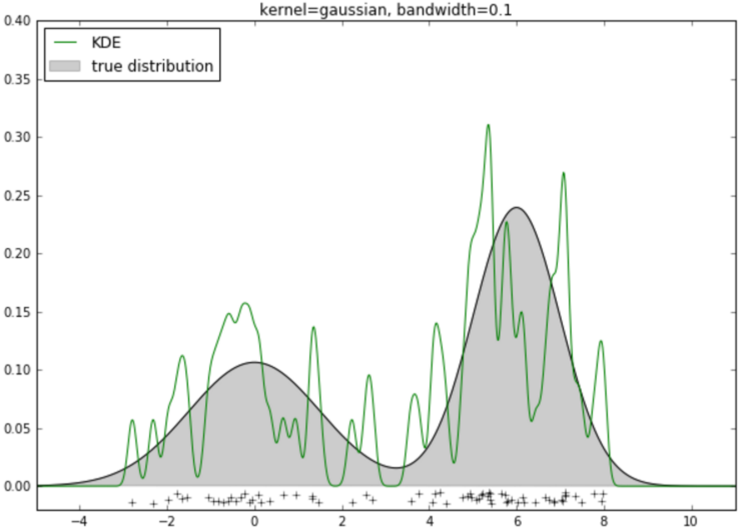
- Результирующая оценка $\widehat{p}(x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right)$ кусочно-постоянна.
- Влияние точки x_i не изменяется с $\rho(x, x_i)$ вблизи x.

Можем использовать гладкие унимодальные ядра[2]:

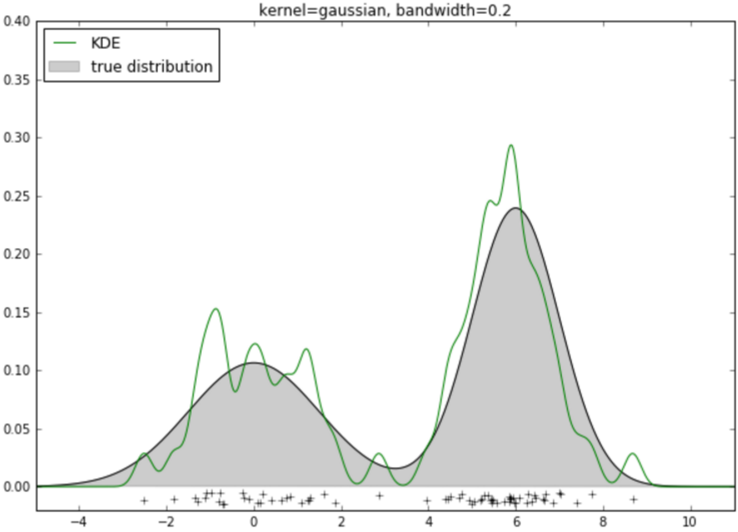


[2] Picture source.

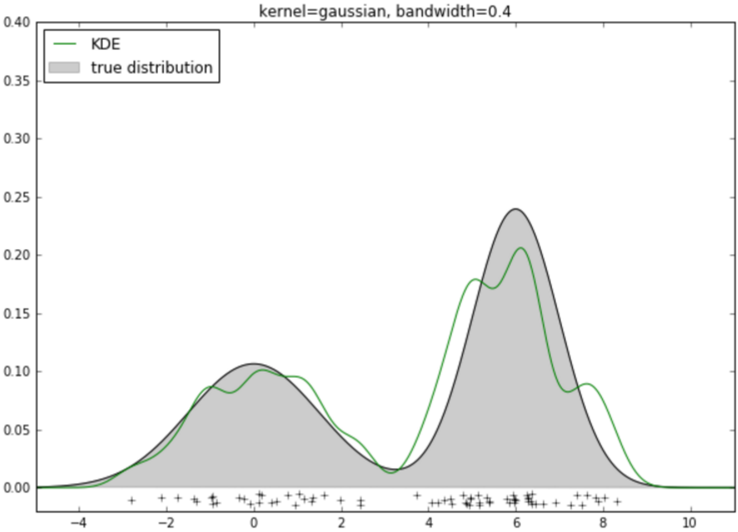
Пример: Гауссово ядро



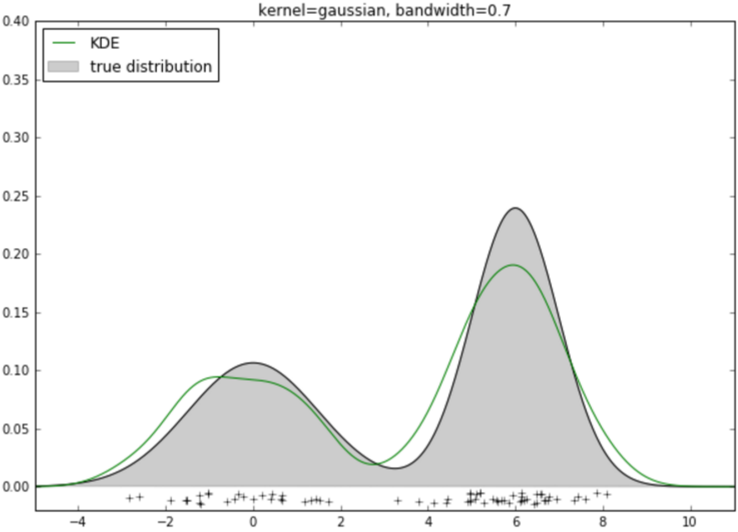
Пример: Гауссово ядро



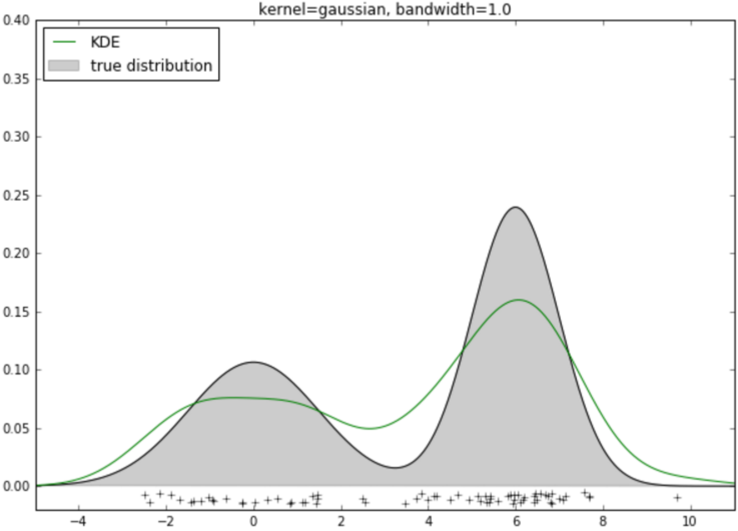
Пример: Гауссово ядро



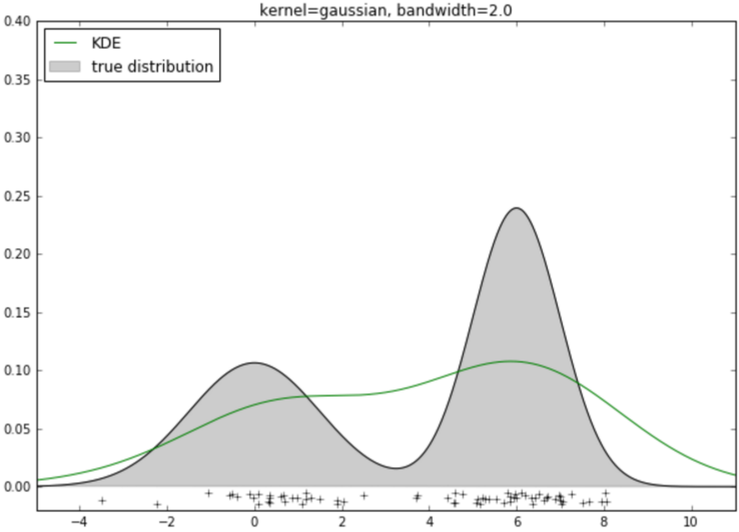
Пример: Гауссово ядро



Пример: Гауссово ядро



Пример: Гауссово ядро



Формулы основных ядер

название	формула $K(u)$
top-hat (прямоугольное)	$\frac{1}{2}\mathbb{I}[\lvert u\rvert \le 1]$
треугольное	$[1-\lvert u\rvert]_+$
Епанечникова	$\propto \left[1-u^2\right]_+$
квартическое	$\propto \left[(1-u^2)\right]_+^2$
Гауссово	$\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$

[3] Как его выбрать?

Формулы основных ядер

название	формула $K(u)$
top-hat (прямоугольное)	$\frac{1}{2}\mathbb{I}[\lvert u\rvert \le 1]$
треугольное	$[1-\lvert u\rvert]_+$
Епанечникова	$\propto \left[1-u^2\right]_+$
квартическое	$\propto \left[(1-u^2)\right]_+^2$
Гауссово	$\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$

Комментарии:

- тип ядра влияет на гладкость, но не на точность аппроксимации.
- для точности важен правильный выбор ширины окна[3].

[3] Как его выбрать?

Условия сходимости оценки

Условия сходимости оценки

Ядерная оценка плотности $\widehat{p}(x)$ сходится $\mathbb{E}[(\widehat{p}(x) - p(x))^2] \xrightarrow{N\to\infty} 0$ для $\forall x$ при выполнении условий ниже.

Условия сходимости оценки

Условия сходимости оценки

Ядерная оценка плотности $\widehat{p}(x)$ сходится $\mathbb{E}[(\widehat{p}(x) - p(x))^2] \xrightarrow{N\to\infty} 0$ для $\forall x$ при выполнении условий ниже.

Достаточные условия сходимости:

- Сходимость ширины окна:
- $\lim_{N\to\infty} h(N) = 0$
- $\lim_{N\to\infty} Nh(N) = \infty$

Условия сходимости оценки

Условия сходимости оценки

Ядерная оценка плотности $\widehat{p}(x)$ сходится $\mathbb{E}[(\widehat{p}(x) - p(x))^2] \xrightarrow{N\to\infty} 0$ для $\forall x$ при выполнении условий ниже.

Достаточные условия сходимости:

- Сходимость ширины окна:
- $\lim_{N\to\infty} h(N) = 0$
- $\lim_{N\to\infty} Nh(N) = \infty$
- Регулярность ядра:
    - $\int |K(u)|du < \infty$
    - $\int K(u)du = 1$
    - $\sup_u K(u) < \infty$
    - $\lim_{u\to\infty} |uK(u)| = 0$

Содержание

Расширение на многомерный случай

Многомерные ядра:

$$\widehat{p}(\mathbf{x}) = \frac{1}{Nh^D} \sum_{i=1}^{N} K\left(\frac{1}{h}(\mathbf{x} - \mathbf{x_i})\right)$$

Для ядра д. быть выполнено: $\int_{-\infty}^{+\infty} ... \int_{-\infty}^{+\infty} K(\mathbf{u}) du_1 ... du_D = 1$

название	формула $K(\mathbf{u})$
Гауссово	$\frac{1}{(2\pi)^{D/2}} e^{-\frac{\mathbf{u}^T\mathbf{u}}{2}}$
Епанечникова	$\propto \left[1 - \mathbf{u}^T\mathbf{u}\right]_+$
произведение одномерных	$\prod_{d=1}^{D} K_{1D}\left(\frac{\mathbf{x}^d - \mathbf{x}_n^d}{h}\right)$

Ядра, зависящие от расстояния

Ядра, основанные на расстоянии:

$$\widehat{p}(\mathbf{x}) = \frac{1}{Nh^D} \sum_{i=1}^{N} K\left(\frac{\rho\left(\mathbf{x}, \mathbf{x}_i\right)}{h}\right)$$

Для ядра д. быть выполнено условие нормировки.

название	формула $K(\rho(x, x_i))$
Гауссово	$\frac{1}{(2\pi)^{D/2}} e^{-\frac{\rho^2(x,x_i)}{2}}$
Епанечникова	$\propto \left[1 - \rho^2(x, x_i)\right]_+$

Выбор ширины окна

Выбор ширины окна

Чем чаще лежат точки, тем меньше должна быть ширина окна h.

[4] Чему будет равна оценка h на обучающей выборке?

Выбор ширины окна

Выбор ширины окна

Чем чаще лежат точки, тем меньше должна быть ширина окна h.

Постоянная ширина окна:

- $h = \frac{1}{N}\sum_{i=1}^{N} d_{iK}$, d_{iK}-расстояние от x_i до K-го ближайшего соседа
- Можно оценить h максимизацией правдоподобия на валидационной выборке[4]

Переменная ширина окна (для изменяющейся частоты точек):

- $h(x)$ = расстояние до K-го ближайшего соседа x

[4] Чему будет равна оценка h на обучающей выборке?

Содержание

Метод Парзеновского окна

Оценим $p(x|y)$ ядерной оценкой плотности:

$$p(x|y) = \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Метод Парзеновского окна

Оценим $p(x|y)$ ядерной оценкой плотности:

$$p(x|y) = \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Байесовское решающее мин. ошибки правило дает **метод Парзеновского окна**:

$$\widehat{y} = \arg\max_y p(y|x) = \arg\max_y p(y)p(x|y)$$

Метод Парзеновского окна

Оценим $p(x|y)$ ядерной оценкой плотности:

$$p(x|y) = \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x,x_i)}{h}\right)$$

Байесовское решающее мин. ошибки правило дает **метод Парзеновского окна**:

$$\widehat{y} = \arg\max_y p(y|x) = \arg\max_y p(y)p(x|y)$$

$$= \arg\max_y \frac{N_y}{N} \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x,x_i)}{h}\right)$$

Метод Парзеновского окна

Оценим $p(x|y)$ ядерной оценкой плотности:

$$p(x|y) = \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Байесовское решающее мин. ошибки правило дает **метод Парзеновского окна**:

$$\widehat{y} = \arg\max_y p(y|x) = \arg\max_y p(y)p(x|y)$$

$$= \arg\max_y \frac{N_y}{N} \frac{1}{N_y h^D} \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

$$\widehat{y}(x) = \arg\max_y \sum_{i:y_i=y} K\left(\frac{\rho(x, x_i)}{h}\right)$$

Метод K-ближайших соседей

- Строим прогноз для x:
  - $K(u) = \mathbb{I}[|u| \le 1]$
  - $h(x)$ = расстояние от x до K-го ближайшего соседа.
- Тогда метод Парзеновского окна дает метод K-ближайших соседей:

$$\widehat{y}(x) = \arg\max_y \sum_{i:y_i=y} K\left(\frac{\rho(x,x_i)}{h(x)}\right)$$

$$= \arg\max_y \sum_{i:\rho(x,x_i)\le h(x)} \mathbb{I}[y_i = y]$$

- Сравнение подходов выбора h:
  - фиксированный h: усредняем по фиксированной окрестности вокруг x
  - адаптивный h (K ближайших соседей): всегда усредняем по $\ge K$ ближайшим точкам

Заключение

- Ядерная оценка плотности - непараметрический метод оценки плотности.
- Выбор ф-ции ядра контролирует непрерывность и дифференцируемость плотности.
- Ширина окна контролирует сложность (гибкость) модели.
- Ширина окна должна быть
  - больше для разреженных областей
  - меньше для густо наполненных областей
- Метод Парзеновского окна и K ближайших соседей - частные случаи правила минимальной ошибки и ядерной оценки плотности.