Politechnika Wrocławska

Regresja jądrowa Nadaraya-Watsona

Jakub Koral

Wrocław, 8 grudnia 2021

HR EXCELLENCE IN RESEARCH

Estymacja gęstości

Rozpiszmy gęstość jako pochodną dystrybuanty

$$\begin{aligned} f(x) = F'(x) &= \lim_{h \to 0} \frac{F(x+h) - F(x-h)}{2h} \\ &= \lim_{h \to 0} \frac{\mathbb{P}[x - h \leqslant X \leqslant x + h]}{2h}. \end{aligned} \tag{1}$$

Teraz wystarczy oszacować prawdopodobieństwo, by otrzymać jądrowy estymator gęstości

$$\begin{aligned} \widehat{f}(x, h) &= \frac{1}{2nh} \sum_{i=1}^{n} \mathbb{1}_{\{x - h \leqslant x_i \leqslant x + h\}} \\ &= \frac{1}{nh} \sum_{i=1}^{n} \frac{1}{2} \cdot \mathbb{1}_{\left\{\left|\frac{x - x_i}{h}\right| \leqslant 1\right\}} = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \end{aligned} \tag{2}$$

Widać, że $K(z) = \frac{1}{2} \cdot \mathbb{1}_{\{|z| \leqslant 1\}} = f_U(z)$, gdzie $U \sim \mathcal{U}(-1, 1)$.

Formalne definicje

Definicja (Jądrowy estymator gęstości)

Jądrowy estymator gęstości dla próbki $\{x_i\}_{i=1}^n$ *zadany jest wzorem*

$$\widehat{f}(x,h) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right),$$

gdzie $K(\cdot)$ *to jądro, a* h *to parametr wygładzenia.*

Definicja (Jądro)

Funkcję $K : \mathbb{R} \longrightarrow [0,\infty)$, *która spełnia następujące warunki*

1. $\int_{\mathbb{R}} K(z)\,\mathrm{d}\,z = 1$ *(normalizacja),*
2. $K(-z) = K(z)$ *(symetria),*

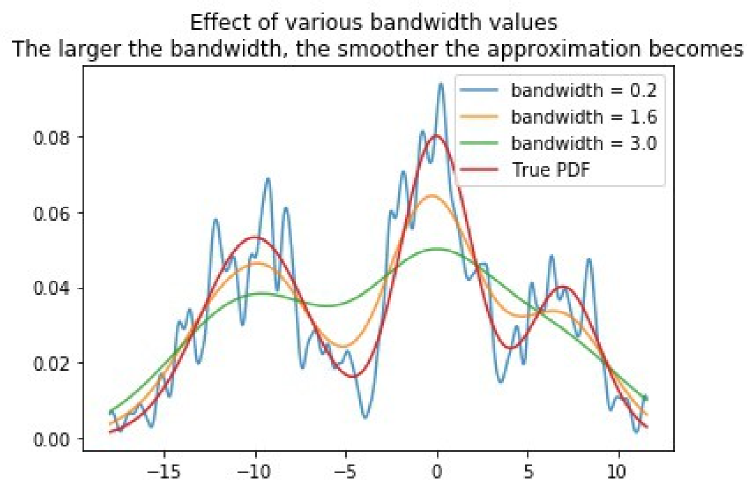
nazywamy jądrem. Często stosujemy zapis $K_h(z) := \frac{1}{h}K\left(\frac{z}{h}\right)$.

Przykłady jąder

Nazwa	Wzór
jądro jednostajne (prostokątne)	$\frac{1}{2} \cdot \mathbb{1}_{\{\lvert z\rvert \leqslant 1\}}$
jądro trójkątne	$(1-\lvert z\rvert) \cdot \mathbb{1}_{\{\lvert z\rvert \leqslant 1\}}$
jądro normalne	$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
jądro Epanecznikowa	$\frac{3}{4}\left(1-z^2\right) \cdot \mathbb{1}_{\{\lvert z\rvert \leqslant 1\}}$
jądro cosinusowe	$\frac{\pi}{4} \cos \frac{\pi z}{2} \cdot \mathbb{1}_{\{\lvert z\rvert \leqslant 1\}}$

Wybór jądra ma niewielkie znaczenie (w porównaniu do wyboru parametru h). Najlepsze pod względem scałkowanego błędu średniokwadratowego jest jądro Epanecznikowa.

Dlaczego parametr wygładzenia jest istotny?



Rysunek: Porównanie estymatorów gęstości jądrowej dla różnych parametrów h (ang. *bandwidth*).

Dobór parametru wygładzenia

Nie da się wyznaczyć optymalnego (w sesnie średniokwadratowym) estymatora parametru h bez znajomości postaci funkcji gęstości f. Można jednak znaleźć dobre jego przybliżenie metodą cross-validation. Zdefiniujmy

$$\mathrm{LSCV}(h) := \int \widehat{f}(x,h)^2 \,\mathrm{d}\,x - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} K_h(x_i - x_j), \tag{3}$$

który nazywamy selektorem cross-validation o najmniejszym błędzie średniokwadratowym. Wtedy możemy oszacować h przy pomocy wzoru

$$\widehat{h}_{\mathrm{LSCV}} := \operatorname*{argmin}_{h>0} \mathrm{LSCV}(h). \tag{4}$$

Regresja nieparametryczna

W przypadku regresji nieparametrycznej będziemy estymować funkcję

$$m(x) = \mathbb{E}[Y|X = x] = \int yf(y|x)\,\mathrm{d}\,y = \int y\frac{f(x,y)}{f(x)}\,\mathrm{d}\,y. \quad (5)$$

Skorzystamy z jądrowego estymatora gęstości

$$\begin{aligned}\widehat{\mathbb{E}}[Y|X = x] &= \int y\frac{\sum_{i=1}^{n} K_h(x - x_i)K_h(y - y_i)}{\sum_{j=1}^{n} K_h(x - x_j)}\,\mathrm{d}\,y \\ &= \frac{\sum_{i=1}^{n} K_h(x - x_i)\int yK_h(y - y_i)\,\mathrm{d}\,y}{\sum_{j=1}^{n} K_h(x - x_j)} \\ &= \frac{\sum_{i=1}^{n} K_h(x - x_i)y_i}{\sum_{j=1}^{n} K_h(x - x_j)}.\end{aligned} \quad (6)$$

Estymator jądrowy Nadaraya-Watsona dany jest wzorem

$$\boxed{\widehat{m}_{NW}(x,h) = \frac{\sum_{i=1}^{n} K_h(x - x_i)y_i}{\sum_{j=1}^{n} K_h(x - x_j)}.} \quad (7)$$

Kod

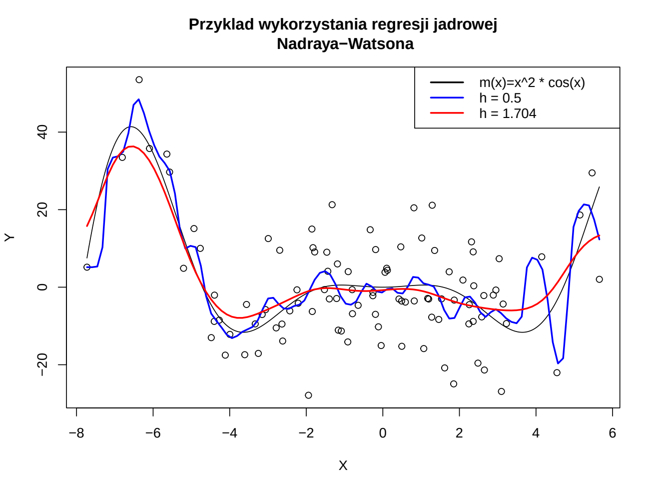
Implementacją regresji jądrowej Nadaraya-Watsona w R jest chociażby funkcja ksmooth:

```
require(graphics)
library(MASS)

n <- 100
X <- rnorm(n, sd = 3)
m <- function(x) x^2 * cos(x)
eps <- rnorm(n, sd = 10)
Y <- m(X) + eps
h_LSCV <- ucv(X, nb = n)
xGrid <- seq(min(X), max(X), l = 500)

plot(X, Y)
lines(xGrid, m(xGrid))
lines(ksmooth(X, Y, "normal", bandwidth = 0.5),
col = "blue", lwd = 2)
lines(ksmooth(X, Y, "normal", bandwidth = h_LSCV),
col = "red", lwd = 2)
title("Przykład wykorzystania regresji jądrowej\n
Nadraya-Watsona")
legend(x="topright", c("m(x)=x^2 * cos(x)","h = 0.5",
paste("h =", toString(round(h_LSCV, 3)))), lty = 1,
col=c("black","blue", "red"), lwd = 2)
```

Przykład



Rysunek: Zależność dla próbki $\{x_i\}_{i=1}^{100}$ z rozkładu $\mathcal{N}(0, 3)$ i $y_i = x_i^2 \cos(x_i) + \varepsilon_i$, gdzie $\varepsilon_i \sim \mathcal{N}(0, 10)$. Dorysowane krzywe regresji jądrowej Nadaraya-Watsona dla parametrów $h = 0.5$ i $h = \widehat{h}_{\mathrm{LSCV}}$.

Źródła

Literatura:

Portugués E.G., *Notes for Predictive Modeling*:

- **Jądrowy estymator gęstości**: https://bookdown.org/egarpor/PM-UC3M/npreg-npdens.html
- **Regresja Nadaraya-Watsona**: https://bookdown.org/egarpor/PM-UC3M/npreg-kre.html

Rysunek:

- https://deepai.org/machine-learning-glossary-and-terms/kernel-density-estimation