Jądrowa estymacja funkcji regresji (*wersja II*)

dr hab. inż. Przemysław Śliwiński, prof. PWr.

4 maja 2017

1 Jądrowy estymator regresji

1. Wygenerować ciąg liczb $\{X_1, \ldots, X_N\}$, $N = 1024$, o dowolnej gęstości w przedziale $[-1, 1]$.
2. Dla uzyskanego ciągu wygenerować wartości
$$Y_n = m(X_n) + Z_n,\ n = 1, \ldots, N$$
gdzie
$$m(x) = \sqrt[3]{x} \text{ oraz } Z_n \sim N(0, 0.05)$$
3. Zaimplementować *jądrowy estymator regresji* o postaci:[1]
$$\bar{m}_h(x) = \frac{\sum_{n=1}^{N} Y_n \cdot K\left(\frac{x-X_n}{h}\right)}{\sum_{n=1}^{N} K\left(\frac{x-X_n}{h}\right)},$$
przyjmując
$$h = 1/2, 1/4, \ldots, 1/128\ \left(= 2^{-1}, 2^{-2}, \ldots, 2^{-7}\right)$$
gdzie $K(x)$ jest jądrem:
   (a) prostokątnym
   $$K(x) = \frac{1}{2} \cdot \mathbf{1}_{[-1,1)}(x),$$
   (b) trójkątnym
   $$K(x) = (1 - |x|) \cdot \mathbf{1}_{[-1,1)}(x),$$
   oraz
   (c) Epanecznikowa
   $$K(x) = \frac{3}{4}\left(1 - x^2\right) \cdot \mathbf{1}_{[-1,1)}(x),$$

[1] ¡Viva Mallorca!

4. Wyznaczyć h takie, aby błąd estymatora, dla każdego z powyższych $K(x)$, dany formułą

$$validation\,(h) = \frac{1}{2Q}\sum_{q=-Q}^{Q}\left[\bar{m}_h\left(\frac{q}{Q}\right) - m\left(\frac{q}{Q}\right)\right]^2$$

   był jak najmniejszy (przyjąć $Q = 100$).

5. Zastąpić zakłócenie normalne Z_n, zakłóceniem o rozkładzie Cauchy'ego $C\,(0, 0.05)$ i powtórzyć eksperyment.

6. Wyciągnąć wnioski.