DATA MINING 2014 Laboratorium 9 (REGRESJA NIEPARAMETRYCZNA)

9.1 (Zadanie domowe)

- Wygeneruj punkty x_i , i = 1, ..., n z rozkładu jednostajnego na [0, 4].
- Wygeneruj $y_i = f(x_i) + \epsilon_i$, gdzie błędy ϵ_i mają rozkład normalny ze średnią 0 i odchyleniem $\sigma = 0.1$.
- Rozważamy 3 funkcje:

$$f(x) = 4.26(e^{-x} - 4e^{-2x} + 3e^{-3x}),$$

$$f(x) = \sin(3x),$$

$$\begin{split} f(x) &= x \cdot 1(x \in (0,1]) + (-x+2) \cdot 1(x \in (1,2]) + \\ (x-2) \cdot 1(x \in (2,3]) + (-x+4) \cdot 1(x \in (3,4]) \end{split}$$

- Dopasuj krzywe używając funkcji: locpoly, ksmooth, smooth.spline, loess na podstawie danych (x_i, y_i) ,
- Oblicz $ISE = n^{-1} \sum_{i=1}^{n} [f(x_i) \hat{f}(x_i)]^2$ dla powyższych funckji, gdzie \hat{f} oznacza dopasowaną funkcję.
- Zbadaj zależność między błędem ISE i liczbą obserwacji $n=50,100,150,\ldots,500.$
- Dla funkcji locpoly and ksmooth sprawdź działanie różnych funkcji jądrowych.
- \bullet Przygotuj wykres pokazujący zależność ISE od n.