Semestrální test 2 - oprava

neděle 15. ledna 2023 15:00

Úloha 1. (2 body)
Mějme jednodimenzionální měření $x \in T = \{-4, 4, -2.5, 0, 3, 0\}$.

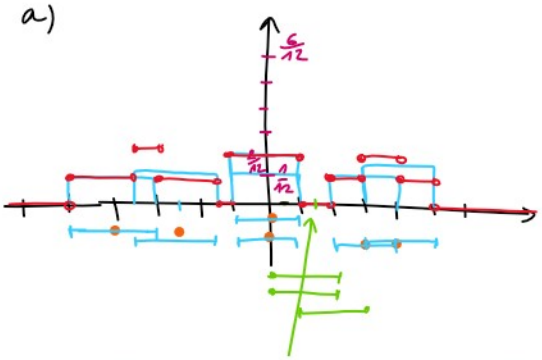
a) Vykreslete graf hustoty rozdělení $\hat{p}(x)$ odhadnuté pomocí Parzenovy metody. Použijte kernel funkci $K(x, y) = k(x - y)$, kde

$$k(z) = \begin{cases} \frac{1}{h}, & \text{pokud } |z| \leq \frac{h}{2} \\ 0, & \text{jinak} \end{cases}$$

a $h = 2$. Správně popište hodnoty na ose y grafu. *(1 bod)*

b) Vypočítejte hodnotu $\hat{p}(x = 1.5)$ pomocí stejného kernelu, ale s $h = 4$. *(1 bod)*

a)



$x \in T = \{-4, 4, -2.5, 0, 3, 0\}$

$K(x,y) = k(x-y) = k(z)$

$k(z) = k(x-y) = \begin{cases} 1/h & |x-y| \leq \frac{h}{2} = 2 \rightarrow |x-y| \leq 1 \\ 0 & \text{d/w} \end{cases}$

Tam, kde je vzdálenost od bodů měření x menší nebo rovna 1, tam je hustota rozdělení 1/2
Všude jinde je hodnota rozdělení nula

$h = 2$, $N = 6$ $\quad y: \frac{1}{h \cdot N} = \frac{1}{2 \cdot 6} = \frac{1}{12}$

b) Vypočítejte hodnotu $\hat{p}(x = 1.5)$ pomocí stejného kernelu, ale s $h = 4$.

$\hat{p}(x=1.5)$ když $h=4 \rightarrow |x-y| = \frac{4}{2} = 2$

$p = \frac{K}{N \cdot V}$ $\quad \frac{3}{N \cdot h} = \frac{3}{6 \cdot 4} = \frac{3}{24} = \frac{1}{8}$

Úloha 2. (2 body)
Rychlost větru je modelována Rayleighovým rozdělením

$$P(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$$

kde σ je parametr a $x \geq 0$. Najděte maximum likelihood odhad $\hat{\sigma}$ z naměřených rychlostí $T = \{0, 1, 2, 2, 1\}$. $x_1, x_2, \ldots, x_5$

$\hat{\sigma} = \arg\max L(\sigma) = \arg\max P(T|\sigma) = \arg\max \prod_{i=1}^{N} p(x_i|\sigma)$

$L(\sigma) = \left(\frac{x_1}{\sigma^2} e^{-\frac{x_1^2}{2\sigma^2}}\right) \cdot \left(\frac{x_2}{\sigma^2} \cdot e^{-\frac{x_2^2}{2\sigma^2}}\right) \cdot \ldots$

$= \frac{x_1 \cdot x_2 \cdot \ldots \cdot x_5}{(\sigma^2)^5} \cdot e^{\frac{-x_1^2}{2\sigma^2} + \frac{-x_2^2}{2\sigma^2} + \ldots}$ $\quad e^{\frac{-x_1^2 - x_2^2 - \ldots - x_5^2}{2\sigma^2}}$

$e^y = x$
$\ln_e x = y$
$\ln_e(e^y) = y$

$\ln L(\sigma) = \prod_{i=1}^{5} \frac{x_i}{\sigma^2} \cdot e^{\frac{-x_i^2}{2\sigma^2}}$

$\log L(\sigma) = l(\sigma) = \ln\left[\prod_{i=1}^{5} \frac{x_i}{\sigma^2} \cdot e^{\frac{-x_i^2}{2\sigma^2}}\right] = \sum_{i=1}^{5} \ln\left(\frac{x_i}{\sigma^2} \cdot e^{\frac{-x_i^2}{2\sigma^2}}\right)$

$= \sum_{i=1}^{5} \ln\left(\frac{x_i}{\sigma^2}\right) + \ln\left(e^{\frac{-x_i^2}{2\sigma^2}}\right) = \sum \ln(x_i) - \ln(\sigma^2) + \underbrace{\ln\left(e^{\frac{-x_i^2}{2\sigma^2}}\right)}_{\frac{-x_i^2}{2\sigma^2}}$

$\frac{\partial l(\sigma)}{\partial \sigma} = \sum_{i=1}^{5} 0 - \frac{1}{\sigma^2} \cdot 2\sigma + (-2) \cdot \frac{-x_i^2}{2\sigma^3} = 0$

$$\frac{\partial \ell(\sigma)}{\partial \sigma} = \sum_{i=1} 0 - \frac{1}{\sigma^2} \cdot 2\sigma + (-2) \cdot \frac{x_i^2}{2\sigma^3} = 0$$

$T = \{0,1,2,2,1\}$ $T = \{0,1,0,0,2\}$ $\sum x_i^2 = 1+4$

$$\frac{\partial \ell(\sigma)}{\partial \sigma} = \sum_{i=1}^{5} -\frac{2\sigma}{\sigma^2} + \frac{2x_i^2}{2\sigma^3} = \sum_{i=1}^{5} -\frac{2}{\sigma} + \frac{x_i^2}{\sigma^3} = 0$$

$$-5 \cdot \frac{2}{\sigma} + \frac{\sum x_i^2}{\sigma^3} = 0$$

$\sum x_i^2 = 1+1+4+4$

$$-\frac{10}{\sigma} + \frac{10}{\sigma^3} = 0$$

$$\frac{-10\sigma^2 + 10}{\sigma^3} = 0$$

$$-10\sigma^2 + 10 = 0$$

$$\sigma^2 = \frac{-10}{-10} = 1$$

$$\hat{\sigma} = \pm\sqrt{1} = \pm 1$$

$$-\frac{10}{\sigma} + \frac{5}{\sigma^3} = 0$$

$$\frac{-10\sigma^2 + 5}{\sigma^3} = 0$$

$$-10\sigma^2 + 5 = 0$$

$$\sigma^2 = \frac{-5}{-10}$$

$$\sigma = \pm\sqrt{\frac{1}{2}}$$

Úloha 3. *(1 bod)* Popište klasifikační algoritmus k-NN.

- k-NN ... k nearest neigborough ... k nejbližších sousedů
- Pro dané x najdu k nejbližších bodů z trénovací množiny - podívám se na jejich klasifikaci a podle nejčetnějšího klasifikuju x