#B241/RPZ

Úloha 1 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 $T = \{1.2, 3.5, 1.1, 0.7, 2.9\}$. *(3 body)*

Úloha 2
Nakreslete příklad trénovacích dat, které základní verze logistické regrese nedokáže správně klasifikovat. Zdůvodněte. *(1 bod)*

Úloha 3
Máme dodat klasifikátor zákazníkovi, který nás odmění na základě chyby klasifikátoru na nám neznámé testovací množině. Máme k dispozici trénovací množinu a umíme natrénovat několik různých klasifikátorů. Jak vybereme jeden, který dodáme zákazníkovi? *(1 bod)*

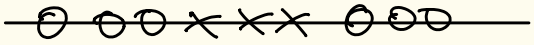
(3) cross validation?

(2)

O - třída o

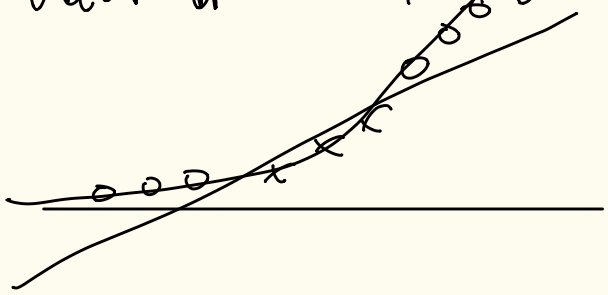
X - třída x

například data



nemůžeme rozdělit základní verzí logistické regrese. Musel bychom

udělat dimensional lifting



(1) $P(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ σ je parametr a $x \geq 0$

$T = \{1.2, 3.5, 1.1, 0.7, 2.9\}$

$L(\sigma) = \prod_i p(t_i)$

$l(\sigma) = \sum_i p(t_i) = \sum_i \left(\log \frac{t_i}{\sigma^2} - \frac{t_i^2}{2\sigma^2} \right)$

$$\frac{\partial l}{\partial \sigma} = \sum_i -2\frac{\sigma^2}{t_i}\cdot\frac{t_i}{\sigma^3} + \frac{t_i^2}{\sigma^3} = \sum_i \frac{-2}{\sigma} + \frac{t_i^2}{\sigma^3} = \frac{-2N}{\sigma} + \sum_i \frac{t_i^2}{\sigma^3}$$

$= 0$ (maximum)

$$\Rightarrow \quad \sum_i \frac{t_i^2}{\sigma^3} = \frac{2N}{\sigma}$$

$$\frac{1}{\sigma^3}\sum_i t_i^2 = \frac{2N}{\sigma}$$

$$\sum_i t_i^2 = 2N\sigma^2$$

$$\sigma = \pm\sqrt{\frac{\sum_i t_i^2}{2N}} = \pm\sqrt{\frac{23{,}8}{10}} = \pm 1{,}54$$

Parzen

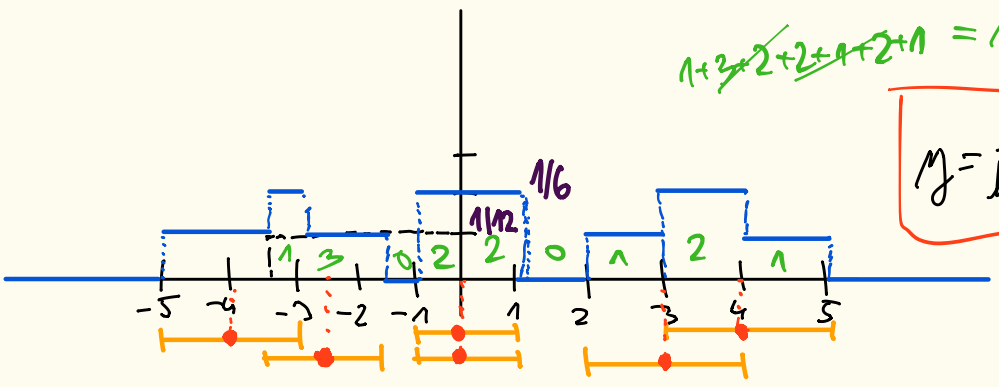
(1) Mějme jednodimenzionální měření $x \in T = \{-4, 4, -2.5, 0, 3, 0\}$

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

$$k(z) = \begin{cases} 1/h, & \text{pokud } |z| \le h/2 \rightarrow |x-y| \le 1 \\ 0, & \text{jinak} \end{cases}$$

a $h = 2$. Správně popište hodnoty na ose y grafu.

(b) vypočítejte hodnotu $\hat{p}(x = 1.3)$ pomocí stejného kernelu, ale s $h = 4$.



$1+3+2+2+1+2+1 = 12$

$$y = \frac{1}{h \cdot N} = \frac{1}{12}$$

(h) pro $h=2$ je $\hat{p}(x=1.5)=0$

pro $h=4$: $|x-y| \le 2$

v 1,5 budou 3 boxy $\Rightarrow \hat{p}(x=1.5) = \frac{3}{4 \cdot 6} = \frac{3}{24} = \frac{1}{8}$

Úloha 2

$$p(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \qquad T = \{0,1,2,2,1\}$$

najděte MLE $\hat{\sigma}$

$$L(\sigma) = \prod_i p(t_i)$$

$$\ell(\sigma) = \sum_i \log \frac{t_i}{\sigma^2} - \sum_i \frac{t_i^2}{2\sigma^2}$$

$$\frac{\partial \ell}{\partial \sigma} = \sum_i \frac{\sigma^2}{t_i} \cdot \left(-2\frac{t_i}{\sigma^3}\right) + \sum_i \frac{t_i^2}{2\sigma^3} =$$

$$= -2 \sum_i \frac{1}{\sigma} + \sum_i \frac{t_i^2}{\sigma^3} =$$

$$= \frac{-2N}{\sigma} + \sum_i \frac{t_i^2}{\sigma^3} \overset{\text{maximum}}{=} 0$$

$$\frac{2N}{\sigma} = \frac{1}{\sigma^3} \sum t_i^2 \quad / \cdot \sigma^3$$

$$2N\sigma^2 = \sum_i t_i^2$$

$$\hat{\sigma} = \pm \sqrt{\frac{\sum_i t_i^2}{2N}} = \pm \sqrt{\frac{10}{2 \cdot 5}} = \pm 1$$

(3) Popište klasifikační algoritmus k-NN

k nejbližších sousedů → z trénovací množiny vyberu k nejbližších, podívám se na jejich klasifikace a dle té nejdetnější klasifikuji daný bod.

Fotbaloví chuligáni – agresivní / klidný fanoušci

agresivní fanoušek ... w počet agresivních A

klidný fanoušek ... $1-w$

N ... celkový počet fanoušků

počet agresivního fanouška

$$w^* = \underset{w}{\operatorname{argmax}} \prod_i p(x_i|w) = w^A \cdot (1-w)^{N-A}$$

$$L(w) = w$$

$$l(w) = A \log w + (N-A)\log(1-w)$$

$$\frac{\partial l}{\partial w} = \frac{A}{w} - \frac{N-A}{1-w} \overset{\text{nast}}{=} 0$$

$$\frac{A}{w} = \frac{N-A}{1-w}$$

$$A(1-w) = (N-A)w$$

$$A - \cancel{Aw} = Nw - \cancel{Aw}$$

$$\hat{w} = \frac{A}{N}$$