

RPZ Test**19. ledna 2021****Jméno:**

1. (6b) Parzenova okna a Neyman-Pearson.

Mějme dvě třídy označené 1 a 2. Trénovací množina pro třídu 1 je $X_1 = \{1, -1, 1, 2, 0\}$, pro třídu 2 je $X_2 = \{-1, 0, 1, 3\}$.

- (a) (3b) Z dat X_1 odhadněte podmíněnou hustotu $p(x|1)$ metodou Parzenových oken. Analogicky odhadněte hustotu $p(x|2)$ pro druhou třídu. Použijte jádrovou funkci $K(x, y) = k(x - y)$, kde $k(z)$ je definovaná takto:

$$\begin{aligned} k(z) &= 1/h \quad \text{pro } |z| \leq h/2, \\ k(z) &= 0 \quad \text{pro } |z| > h/2, \end{aligned}$$

kde $h = 2$. Odhadnuté hustoty graficky znázorněte. Spočítejte $p(x|1)$ a $p(x|2)$ pro $x = 0.5$.

- (b) (3b) Třída 2 představuje nebezpečí. Zkonstruujte Neyman-Pearsonovu strategii takovou, aby přehlédnuté nebezpečí bylo < 40%. Pro nalezenou strategii spočítejte chyby klasifikace ϵ_1, ϵ_2 pro obě třídy 1 a 2.

2. (8b) Adaboost.

- (a) (3b) Vysvětlete metodu Adaboost tak, aby ji podle Vašeho popisu kterýkoli Váš kolega dokázal naimplementovat.
- (b) (1b) Popište časovou náročnost učení.
- (c) (1b) Diskutujte časovou náročnost klasifikace.
- (d) (1b) Napište, jestli a jak je možné chod algoritmu zrychlit při učení/klasifikaci.
- (e) (2b) Popište výhody Adaboostu ve srovnání s SVM, logistickou regresí, nejbližším sousedem a neuronovými sítěmi (zvažte jak fázi učení, tak i fázi klasifikace.)

3. (3b) Maximálně věrohodný odhad parametrů. Normální rozdělení je dáno jako

$$p(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} \det(\mathbf{C})^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})},$$

kde $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^d$, $\mathbf{C} \in \mathbb{R}^{d \times d}$, $\mathbf{C} \succeq 0$. Mějme N nezávislých pozorování $(x_i)_{i=1}^N$ z tohoto rozdělení. Kovarianční matice \mathbf{C} je známá. Najděte maximálně věrohodný odhad parametru $\boldsymbol{\mu}$.

4. (3b) Výběr klasifikátoru.

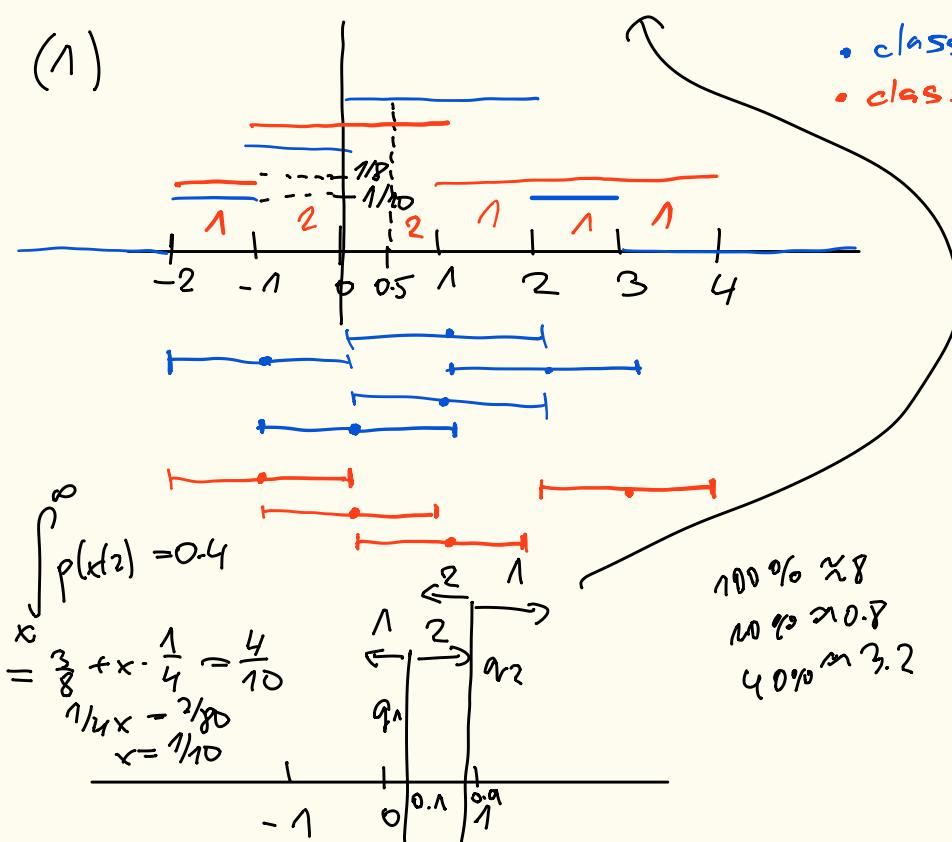
Máte k dispozici trénovací data a máte za úkol implementovat jednoduchý klasifikátor. Chcete, aby klasifikátor fungoval co nejlépe. Rozhodl/a jste se, že začnete s metodou K -nejbližšího souseda (K -NN). Vysvětlete, jak vyberete K .

Jan ukazuje na S 2 = 60%

$$q_1(x) = \begin{cases} 1 & x < 0.1 \\ 2 & \text{else} \end{cases} \quad \epsilon_1 = 67\%$$

$$q_2(x) = \begin{cases} 1 & x > 0.9 \\ 2 & \text{else} \end{cases} \quad \epsilon_2 = 57\%$$

(1)



- class 1
- class 2

$$f(x-y) \quad h=2$$

$$\frac{1}{2} \text{ mo } |z| \leq 1$$

$$0 \text{ mo } |z| > 1$$

$$y_1 = \frac{1}{h \cdot N} = \frac{1}{10}$$

$$\vee 0.5 \text{ je } p(x|1) = \underline{\underline{\frac{3}{10}}}$$

$$y_2 = \frac{1}{h \cdot 4} = \frac{1}{8}$$

$$\vee 0.5 \text{ je } p(x|2) = \underline{\underline{\frac{1}{4}}}$$

(3)

$$p(x) = (2\pi)^{-d/2} \det(C)^{-1/2} e^{-1/2(x-\mu)^T C^{-1}(x-\mu)}$$

$$L(\mu) = \prod_{x \in X} p(x)$$

$$l(\mu) = \sum_{x \in X} -\frac{d}{2} \log(2\pi) - \frac{1}{2} \log(\det(C)) - \frac{1}{2} (x-\mu)^T C^{-1} (x-\mu)$$

$$l(\mu) = -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\det(C)) - \frac{1}{2} \sum_{x \in X} (x-\mu)^T C^{-1} (x-\mu)$$

$$\frac{\partial l}{\partial \mu} = 0 + 0 - \frac{1}{2} \sum_{x \in X} (x-\mu)^T C^{-1} (x-\mu)$$

$$w = -\frac{1}{2} \underbrace{(x-\mu)^T}_{2} \underbrace{C^{-1}}_{A} \underbrace{(x-\mu)}_{3}$$

$$\frac{\partial l}{\partial \mu} = \sum_{x \in X} \frac{\partial w}{\partial \mu}$$

$$w = -\frac{1}{2} y^T A y \rightarrow \frac{\partial w}{\partial \mu} = y^T (A + A^T)$$

