

## Ex. Partial 2 --- Barem de notare

-----

### Problema 1: 3.5p

#### AdaBoost

**a: 2.7p**, defalcăt astfel:

0.6p = 3 x 0.2p (cite 0.2p pt. D1, D2 și respectiv D3)

1.2p = 6 x 0.2p (cite 0.2p pentru cele 6 justificări (vedeti ce am încercuit pe prima pagină a rezolvărilor), sau raționamente / calcule echivalente)

0.6p = 3 x 0.2p (cite 0.2p pt. fiecare dintre  $\epsilon_1$ ,  $\epsilon_2$ ,  $\epsilon_3$ )

0.3p = 3 x 0.1p (cite 0.1p pt. fiecare dintre  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ )

**b: 0.8p**, defalcăt astfel:

0.3p:  $\text{err}(H_3)$

0.3p: zonele de decizie (defalcăt astfel: 0.2p pentru indicarea zonelor, iar 0.1p pentru justificare - cum au fost deduse semnele acestor zone)

0.2p: granițele de decizie

### Problema 2: 1.75p

#### Clusterizare ierarhică

0.6p pentru iterațiile 1 și 2 (adică identificarea corectă a clusterelor C1 și C2)  
= 3 x 0.2 (cite 0.2p pentru fiecare dintre cele 3 similarități: single-, complete- și average-linkage)

0.6p pentru scrierea corectă a condițiilor / justificărilor \*1, \*2, \*3  
= 3 x 0.2p (cite 0.2p pt. SL, CL și AL), iar acest 0.2 se descompune astfel: 0.2 = 4 x 0.05, fiindcă sunt de calculat / evaluat 3 distanțe, plus de identificat care este cea minimă dintre ele)

0.3p = 3 x 0.1p pt. identificarea corectă a clusterului C3, în fiecare dintre cele 3 cazuri: SL, CL și AL.

0.25p pentru concluzie: doar CL ajunge să producă (după 3 iterații) ierarhia aplatizată din eununt.

### Problema 3: 1.75p

#### K-means: criteriul J

**a: 0.5p** = 0.2p (indicare  $\alpha = 0$ ,  $\beta = n$ ) + 0.2p (justificarea valorilor lui  $\alpha$  și  $\beta$ ) + 0.1p (pentru indicarea  $\alpha + \beta = 100$  când  $n = 100$ )

**b: 1.25p** = 0.2p (reprezentarea grafică) + 0.8p (raționament) + 0.25p (calculul lui  $\alpha = 1/2$ )

### Problema 4: 1.25p

#### Distributia Bernoulli -- MLE

**a. 1.05p** = 7 x 0.15p (cite 0.15p pentru fiecare dintre cei 7 pași ai demonstrației)

**b. 0.2p**

### Problema 5: 2.85p (plus eventual 0.5p -- pentru reprezentarea grafică a modelului probabilist)

#### AdaBoost

**a.**

**i: 0.35p** = 0.15p + 2 x 0.1p pentru indicarea corectă a datelor / variabilelor observabile, în genul următor (sau echivalent):  $X_i = (X_{i,1}, \dots, X_{i,10})$ ,  $(X_{i,j} \mid Z=1) \sim \text{Bernoulli}(\theta_A)$  și  $(X_{i,j} \mid Z=0) \sim \text{Bernoulli}(\theta_B)$

**ii: 0.2p** pentru indicarea corectă a variabilelor neobservabile

**iii: 0.3p** = 3 x 0.1p pentru indicarea parametrilor  $\pi$ ,  $\theta_A$  și  $\theta_B$

**0.5p** = 5 x 0.1p pentru indicarea cuvintelor-cheie: (1) mixtură, (2) vector de..., (3) distribuții Bernoulli, (4) independente, (5) identic distribuite (relativ la alegerea monedei)

**Bonus: 0.5p** pentru reprezentarea grafica a modelului (vedeti versiunea noua a problemei 7, de pe site-ul Piazza)

**b.**

**1.5p** =  $6 \times 0.25$  (cite 0.25p pentru fiecare dintre cei 6 pasi ai deducerii functiei de log-verosimilitate a datelor complete la iteratia  $t$ )

Acest 0.25p se descompune astfel: 0.2p pentru calculul corect la pasul respectiv si 0.05p pentru justificarea egalitatii respective