Învăţare automată

— Licenţă, anul III, 2020-2021, testul 1 —

Nume student:

1. **(Algoritmul ID3: aplicare când există instanţe duplicate)**

Setul de date (ipotetice) alăturat va fi folosit pentru a decide dacă un student este harnic (H) sau leneş (L), folosind următoarele atribute: ponderalitate (N/U), culoarea ochilor (A/V) şi numărul de examene promovate în ultima sesiune (2, 3 sau 4).

Ponderalit.	CulOchi	ExPromov	Output
N	A	2	L
N	V	2	L
N	V	2	L
U	V	3	L
U	V	3	L
U	A	4	H
N	A	4	H
N	V	4	H
U	A	3	H
U	A	3	H

a. Ce atribut va alege algoritmul ID3 pentru nodul rădăcină? Jusificaţi riguros.

b. Elaboraţi în întregime arborele produs de ID3 (fără pruning) pentru aceste date.

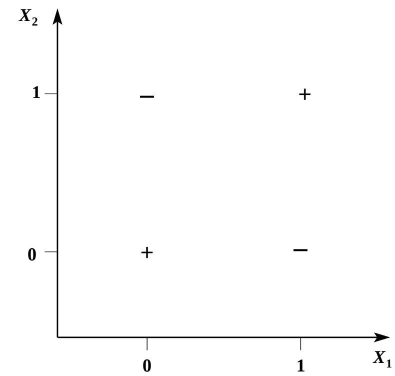
c. Care este eroarea la antrenare pentru arborele ID3?

Observaţie: Pentru calcule puteţi folosi aproximările următoare pentru entropia distribuţiei Bernoulli: $H(1/5) = 0.97095$, $H(2/5) = 0.72193$.

2. **(Comparaţie: algoritmul ID3 (cu pruning incorporat) şi algoritmul Bayes Naiv; aplicare pe date de tip ($\neg$)XOR; comparaţie şi cu regresia logistică)**

• ○ ***CMU, 2011 spring, Tom Mitchell, midterm, pr. 3***

Considerăm o problemă de clasificare cu două variabile booleene $X_1, X_2 \in \{0, 1\}$ şi eticheta $Y \in \{0, 1\}$. În figura alăturată avem două exemple pozitive (+) şi două exemple negative (−). (Observaţie: atât X_1 cât şi X_2 trebuie văzute ca variabile discrete de tip Bernoulli.)



a. Desenaţi un arbore de decizie care poate clasifica în mod perfect aceste patru exemple.

b. Presupunem că folosim următorul algoritm [versiune a lui ID3, cu pruning] pentru elaborarea unui arbore de decizie: începem cu un singur nod (rădăcina arborelui) şi partiţionăm în mod iterativ mulţimea de instanţe asociate fiecărui nod cu ajutorul celui mai „bun“ atribut, care este selectat prin maximizarea câştigului de informaţie la partiţionare; ne vom opri atunci când 1) exemplele asociate unui nod sunt toate de acelaşi tip / clasă, sau 2) nu mai putem găsi niciun atribut care să producă prin partiţionare un câştig de informaţie *strict* pozitiv.

Dacă aplicăm acest algoritm asupra celor patru exemple date, vom obţine oare un arbore de decizie care clasifică în mod perfect aceste exemple? Explicaţi pe scurt ce se va întâmpla.

c. Presupunem că învăţăm un clasificator Bayes Naiv bazat pe exemplele din figura dată, folosind metoda de estimare a propbabilităţilor în sensul verosimilităţii maxime (engl., maximum likelihood estimation, MLE). Listaţi toţi parametrii algoritmului, precum şi valorile estimate asociate lor. Va cataloga oare în mod perfect acest clasificator cele patru exemple?

d. Există oare vreun clasificator de tip regresie logistică bazat pe trăsăturile X_1 şi X_2, capabil să clasifice în mod perfect exemplele din figura dată? De ce da, sau de ce nu?

3. **(Algoritmul Bayes Naiv: calculul ratei medii a erorii – exemplificare)**

Consider a binary classification problem with variable $X_1 \in \{0,1\}$ and label $Y \in \{0,1\}$. The true generative distribution $P(X_1, Y) = P(Y)P(X_1|Y)$ is shown in the following tables:

Y	0	1
$P(Y)$	0.8	0.2

$P(X_1\|Y)$	$Y=0$	$Y=1$
$X_1 = 0$	0.7	0.3
$X_1 = 1$	0.3	0.7

a. Suppose that we have trained a Naive Bayes classifier, using infinite *training data* generated according to the above tables. Please write down in the following table the *predictions* from the trained Naive Bayes for different configurations of X_1. Note that $\hat{Y}(X_1)$ in the table is the decision about the value of Y given X_1. For decision terms in the table, write down either $\hat{Y} = 0$ or $\hat{Y} = 1$; for probability terms in the table, write down the actual values (and the calculation process, e.g., $0.8 \cdot 0.7 = 0.56$).

	$P(X_1, Y=0)$	$P(X_1, Y=1)$	$\hat{Y}(X_1)$
$X_1 = 0$			
$X_1 = 1$			

b. What is the *expected error rate* of this Naive Bayes classifier on *testing examples* that are generated according to the given (first two) tables? In other words, [calculate] $P(\hat{Y}(X_1) \neq Y)$ when (X_1, Y) is generated according to the two tables.

For the next three questions, consider two variables X_1, $X_2 \in \{0,1\}$ and the label $Y \in \{0,1\}$. Y and X_1 are still generated according to the given (first two) tables, and then X_2 is created as a *duplicated copy* of X_1.

c. Suppose now that we have trained a Naive Bayes classifier, using infinite *training data* that are generated according to the given (first two) tables and the duplication rule. In the following table, please write down the *predictions* from the trained Naive Bayes for different configurations of (X_1, X_2).

	$\hat{P}(X_1, X_2, Y=0)$	$\hat{P}(X_1, X_2, Y=1)$	$\hat{Y}(X_1, X_2)$
$X_1 = 0, X_2 = 0$			
$X_1 = 0, X_2 = 1$			
$X_1 = 1, X_2 = 0$			
$X_1 = 1, X_2 = 1$			

d. What is the *expected error rate* of this Naive Bayes classifier on testing examples that are generated according to the given (first two) tables and the duplication rule?

e. Compared to the scenario without X_2, how does the expected error rate change (i.e., increase or decrease)? In the previous table (at point c), the decision rule $\hat{Y}$ on *which* configuration is responsible to this change? What actually happened to this decision rule? (You need to *briefly* answer: increase or decrease, the responsible configuration, and what happened.)

4. **(„Blestemul marilor dimensiuni“)**

In this question, we will try to mathematically characterize what curse of dimensionality means. Specifically, let us consider a d-sphere (just as sphere is 3-dimensional generalization of a circle, d-sphere is a d-dimensional generalization of a circle). The volume of a d-sphere with radius r is given by,

$$V_d(r) = \frac{2\pi^{d/2}}{d\,\Gamma(d/2)} r^d = C_d\, r^d,$$

where C_d is some constant which depends on dimensionality.

To understand the curse of dimensionality, we first compute the fraction of volume lying near the boundaries of the sphere.

a. Compute the expression $V_d(1)$, the volume a d-sphere with radius 1.

b. Compute the expression $V_d(1-\varepsilon)$, the volume of a d-sphere with radius $1-\varepsilon$.

c. Now compute the fraction

$$frac_d(\varepsilon) = \frac{V_d(1) - V_d(1-\varepsilon)}{V_d(1)}$$

d. Plot this fraction $frac_d(\varepsilon)$ with $\varepsilon = 1/2$ for different values of $d = 1, 2, 4, 10$ on the x-axis.

e. What do you observe as d increases? What can you say about the fraction of volume near the boundaries?