Notițe Seminar 6

November 11, 2019

Intro: Până acum am făcut 2 algoritmi de clasificare: ID3, AdaBoost. Mai adăugam acum alți doi algoritmi: **Bayes Naiv** și **Bayes Corelat**.

Remember 1

Fie funcția $f : \{1, 2, 3\} \to \mathbb{R}$ cu $f(1) = 100, f(2) = 200, f(3) = 300$. Atunci:

$$\max_{x \in \{1,2,3\}} f(x) = 300$$

și

$$\arg\max_{x \in \{1,2,3\}} f(x) = 3$$

Remember 2

Ex.: Estimați în sensul verosimilății maxime (MLE) probabilitățile $P(A = 0, B = 0)$, $P(A = 0|Y = 1)$, având următorul tabel cu date:

A	B	Y
0	0	0
0	1	1
1	0	1
1	1	0

Vom nota cu #(proprietăți) câte rânduri au anumite proprietăți. De exemplu: $\#(A = 0) =$ (câte rânduri au $A = 0$).

$P(A = 0, B = 0) \overset{\text{MLE}}{=} \frac{\#(A=0,B=0)}{4} = \frac{1}{4}$

$P(A = 0|Y = 1) \overset{\text{MLE}}{=} \frac{\#(A=0,Y=1)}{\#(Y=1)} = \frac{1}{2}$ (ne uităm doar la rândurile cu $Y = 1$)

Înspre clasificare bayesiană

5.

Classes of Hypotheses

Maximum Likelihood (ML) hypothesis:

the hypothesis that best explains the training data

$$h_{ML} = \operatorname*{argmax}_{h_i \in H} P(D|h_i)$$

Maximum A posteriori Probability (MAP) hypothesis:

the most probable hypothesis given the training data

$$h_{MAP} = \operatorname*{argmax}_{h \in H} P(h|D) = \operatorname*{argmax}_{h \in H} \frac{P(D|h)P(h)}{P(D)} = \operatorname*{argmax}_{h \in H} P(D|h)P(h)$$

Note: If $P(h_i) = P(h_j), \forall i, j$, then $h_{MAP} = h_{ML}$

(slide preluat din `https://profs.info.uaic.ro/~ciortuz/SLIDES/ml6.pdf`)

Aplicare: Vezi ex. 3/pag. 370.

Clasificare bayesiană
Bayes Naiv

11.

2.4 The Naive Bayes Classifier

When to use it:

- **The target function f takes value from a finite set $V = \{v_1, \ldots, v_k\}$**
- **Moderate or large training data set is available**
- **The attributes $< a_1, \ldots, a_n >$ that describe instances are conditionally independent w.r.t. to the given classification:**

$$P(a_1, a_2 \ldots a_n | v_j) = \prod_i P(a_i | v_j)$$

The most probable value of $f(x)$ is:

$$\begin{aligned} v_{MAP} &= \operatorname*{argmax}_{v_j \in V} P(v_j | a_1, a_2 \ldots a_n) = \operatorname*{argmax}_{v_j \in V} \frac{P(a_1, a_2 \ldots a_n | v_j) P(v_j)}{P(a_1, a_2 \ldots a_n)} \\ &= \operatorname*{argmax}_{v_j \in V} P(a_1, a_2 \ldots a_n | v_j) P(v_j) = \operatorname*{argmax}_{v_j \in V} \prod_i P(a_i | v_j) P(v_j) \stackrel{not.}{=} v_{NB} \end{aligned}$$

This is the so-called *decision rule* of the Naive Bayes classifier.

(slide preluat din `https://profs.info.uaic.ro/~ciortuz/SLIDES/ml6.pdf`)

Atribute de intrare: $A_1, \ldots, A_n$

Atributul de ieșire: V

Antrenare: estimarea/calculul următoarelor probabilități:

$P(v_j), \forall v_j \in \mathrm{Val}(V)$

$P(a_i|v_j), \forall a_i \in \mathrm{Val}(A_i), \forall v_j \in \mathrm{Val}(V)$

Număr de parametri necesari de estimat: vezi ex. indicat mai jos

Testare: utilizați regula de decizie din slide (v_{NB}).

Algoritmul este *naiv*, pentru că face presupunerea că atributele de intrare sunt independente condițional față de atributul de ieșire, lucru care nu se întâmplă de cele mai multe ori.

Bayes Corelat

12.

The Joint Bayes Classifier

$$
\begin{aligned}
v_{MAP} &= \operatorname*{argmax}_{v_j \in V} P(v_j | a_1, a_2 \ldots a_n) = \ldots \\
&= \operatorname*{argmax}_{v_j \in V} P(a_1, a_2 \ldots a_n | v_j) P(v_j) = \operatorname*{argmax}_{v_j \in V} P(a_1, a_2 \ldots a_n, v_j) \stackrel{not.}{=} v_{JB}
\end{aligned}
$$

(slide preluat din `https://profs.info.uaic.ro/~ciortuz/SLIDES/ml6.pdf`)

Atribute de intrare: $A_1, \ldots, A_n$

Atributul de ieșire: V

Antrenare: estimarea/calculul următoarelor probabilități:

$P(v_j), \forall v_j \in \text{Val}(V)$

$P(a_1, \ldots, a_n | v_j), \forall a_i \in \text{Val}(A_i), \forall v_j \in \text{Val}(V)$

Număr de parametri necesari de estimat: vezi ex. indicat mai jos

Testare: utilizați regula de decizie din slide (v_{JB}).

Aplicare + altele (număr de parametri de estimat): vezi ex. 7/pag. 379

O altă perspectivă asupra algoritmilor Bayes naiv și Bayes corelat

Regula de decizie a ambilor algoritmi poate fi scrisă și astfel:

$\arg\max_{v_j \in \mathrm{Val}(V)} P_{NB/JB}(a_1, \ldots, a_n, v_j)$

doar că unul calculează probabilitatea folosindu-se de distribuția presupusă de el (mă refer la Bayes naiv care presupune independența condițională a atributelor de intrare față de ieșire), iar altul calculează probabilitatea folosindu-se de distribuția reală.

Altfel spus: având un rând la testare $(a_1, \ldots, a_n)$, gândiți-vă la rândurile $(a_1, \ldots, a_n, v_1)$, $(a_1, \ldots, a_n, v_2)$, ..., $(a_1, \ldots, a_n, v_k)$. Calculați probabilitatea fiecărui rând și returnați eticheta corespunzătoare probabilității maxime.

Netezirea Laplace

În cadrul algoritmului Bayes Naiv, când vreuna din probabilitățile $P(a_i|v_j)$ este 0, va fi o problemă pentru că produsul care apare în regula de decizie a lui Bayes Naiv va fi 0. Pentru a scăpa de acest neajuns, se folosește regula lui Laplace (netezirea de tip *add-one*):

$$P(A_i = a_i | V = v_j) \stackrel{\text{Laplace}}{=} \frac{\#(A_i = a_i, V = v_j) + 1}{\#(V = v_j) + |\mathrm{Val}(A_i)|}$$

Vă amintesc că fără regula lui Laplace era astfel:

$$P(A_i = a_i | V = v_j) \stackrel{\text{MLE}}{=} \frac{\#(A_i = a_i, V = v_j)}{\#(V = v_j)}$$

Aplicare: vezi ex. 6/pag. 377

Dacă vreți să înțelegeți de unde vine formula, citiți în continuare:

Suntem în contextul ex. 6b/pag. 377 și vrem să calculăm $P(\text{study} = 1|\text{category} = \text{spam})$, $P(\text{study} = 1|\text{category} = \text{regular})$ cu regula lui Laplace.

Vom construi un tabel cu frecvențe (valoarea din prima celulă va fi egală cu numărul de rânduri cu study = 0 și category = regular):

	category = regular	category = spam
study = 0	1	8
study = 1	3	0
	4	8

Astfel, am putea scrie:

$$P(\text{study} = 1|\text{category} = \text{regular}) \stackrel{\text{MLE}}{=} \frac{3}{4}$$

$$P(\text{study} = 1 | \text{category} = \text{spam}) \overset{\text{MLE}}{=} \frac{0}{8}$$

Pentru că avem o celulă cu zero, atunci refacem tabelul adăugând 1 (*add-one...*) la fiecare celulă:

	category = regular	category = spam
study = 0	2	9
study = 1	4	1
	6	10

Astfel, putem scrie:

$$P(\text{study} = 1 | \text{category} = \text{regular}) \overset{\text{Laplace}}{=} \frac{4}{6}$$

$$P(\text{study} = 1 | \text{category} = \text{spam}) \overset{\text{Laplace}}{=} \frac{1}{10}$$

Schemă de final

1. Ipoteze
   (a) ML
   (b) MAP
2. Bayes Naiv
   (a) Antrenare
      i. algoritm
      ii. număr de parametri de estimat
   (b) Testare
3. Bayes Corelat
   (a) Antrenare
      i. algoritm
      ii. număr de parametri de estimat
   (b) Testare
4. O altă perspectivă asupra alg. BN, JB
5. Netezirea Laplace