Ex: 33

	T	F
P(Hike)	0.5	0.5

	P(S|H)	P(S|¬H)
Sunny = T	0.8	0.7
Sunny = F	0.2	0.3

	P(W|H)	P(W|¬H)
Windy = T	0.4	0.5
Windy = F	0.6	0.5

33. **(Algoritmului Bayes Naiv: calculul ratei medii a erorilor)**

■ • ∘ *CMU, 2010 fall, Aarti Singh, HW1, pr. 4.2*

Considerăm următoarea problemă de clasificare:
Fie variabila aleatoare Y : *Hike* $\in \{T, F\}$ care denotă faptul că Alice şi Bob merg sau nu în drumeţie în funcţie de condiţiile vremii: X_1 : *Sunny* $\in \{T, F\}$ şi X_2 : *Windy* $\in \{T, F\}$.

Se presupune că au fost estimaţi următorii parametri:

$$P(\textit{Hike}) = 0.5$$
$$P(\textit{Sunny} \mid \textit{Hike}) = 0.8, \quad P(\textit{Sunny} \mid \neg\textit{Hike}) = 0.7$$
$$P(\textit{Windy} \mid \textit{Hike}) = 0.4, \quad P(\textit{Windy} \mid \neg\textit{Hike}) = 0.5$$

De asemenea, se consideră că este satisfăcută presupoziţia de independenţă condiţională a algoritmului Bayes Naiv.

a. Care este probabilitatea (comună) ca Alice şi Bob să meargă în drumeţie atunci când vremea este însorită şi bate vântul, adică

$$P(\textit{Sunny} = T, \textit{Windy} = T, \textit{Hike} = T) = ?$$

Care este decizia luată de algoritmul Bayes Naiv în acest caz?

$$y_{MAP} = \underset{y \in \{T,F\}}{\operatorname{argmax}} P(H = y \mid S = T, W = T) = \underset{y \in \{T,F\}}{\operatorname{argmax}} \frac{P(S=T, W=T \mid H=y)\, H=y)}{P(S=T, W=T, H=T)} =$$

constantă

$$y_{NB} = \underset{y \in \{T,F\}}{\operatorname{argmax}} P(S=T, W=T \mid H=y)$$

$$p_0 \quad P(S=T, W=T, H=T) = P(S=T \mid H=T) \cdot P(W=T \mid H=T) \cdot P(H=T) =$$
$$= 0.8 \cdot 0.4 \cdot 0.5 = 0.40 \cdot 0.4 = 0.16.$$

! Care este prob. cu care va lua decizia? $\frac{p_0}{p_0 + p_1}$

$$\frac{0.16}{0.16 + 0.175} = \frac{0.16}{0.335} = 0.477612$$

b. Completaţi tabelul următor:

X_1	X_2	Y	$P(X_1, X_2, Y)$	$P_{NB}(Y \mid X_1, X_2)$	*decizia algoritmului Bayes Naiv*
F	F	F	0.075	0.555556	F
F	F	T	0.060	0.444444	F
F	T	F	0.075	0.652174	F
F	T	T	0.040	0.347826	F
T	F	F	0.175	0.421686	T
T	F	T	0.240	0.578314	T
T	T	F	0.175	1-0.477612=0.522388	F
T	T	T	0.16	0.477612	F

Observaţie: **Calculele de la punctul *a* corespund ultimei linii din tabelul de mai sus.**

$P(X_1=T \mid Y=T) \cdot P(X_2=F \mid Y=T) \cdot P(Y=T) = 0.8 \cdot 0.6 \cdot 0.5 = 0.240$

$P(X_1=T \mid Y=F)\, P(X_2=F \mid Y=F) \cdot P(Y=F) = 0.7 \cdot 0.5 \cdot 0.5 = 0.175$

$P(X_1=F \mid Y=T) \cdot P(X_2=T \mid Y=T) \cdot P(Y=T) = 0.2 \cdot 0.4 \cdot 0.5 = 0.040$

$P(X_1=F \mid Y=F)\, P(X_2=T \mid Y=F) \cdot P(Y=F) = 0.3 \cdot 0.5 \cdot 0.5 = 0.075$

$P(X_1=F \mid Y=T) \cdot P(X_2=F \mid Y=T) \cdot P(Y=T) = 0.2 \cdot 0.6 \cdot 0.5 = 0.060$

$P(X_1=F \mid Y=F)\, P(X_2=F \mid Y=F) \cdot P(Y=F) = 0.3 \cdot 0.5 \cdot 0.5 = 0.025$

$P_{NB}(Y=T \mid X_1=T, X_2=F) = \frac{p_0}{p_0+p_1} = \frac{0.240}{0.240+0.175} = 0.578314$

⋮

c)

c. Care este rata medie a erorilor (engl., expected error rate) produse de algoritmul Bayes Naiv? Vă reamintim că această (rată) medie este definită ca fiind suma probabilităţilor $P(X_1, X_2, Y)$ pentru acele (triplete de) valori ale variabilelor X_1, X_2, Y pentru care decizia luată de algoritmul Bayes Naiv diferă de valoarea variabilei Y.

În cele ce urmează se presupune că se obţin mai multe informaţii despre vreme. Se introduce o nouă trăsătură X_3: *Rainy* $\in \{T, F\}$. Se presupune că în fiecare zi vremea poate fi fie *Rainy* fie *Sunny*, dar nu şi *Rainy* şi *Sunny*. Similar, se presupune că vremea nu poate fi într-o zi $\neg$*Rainy* şi $\neg$*Sunny*.

$err \overset{def}{=} E_P[1_{\{Y_{NB}(x_1,x_2) \neq Y\}}]$

$= 0.160 + 0.175 + 0.040 + 0.060 = 0.435$

d. În noile condiţii, presupoziţia de independenţă condiţională rămâne oare adevărată? Justificaţi.

e. Calculaţi $P(\textbf{\textit{Sunny}} = T, \textbf{\textit{Windy}} = T, \textbf{\textit{Rainy}} = F, \textbf{\textit{Hike}} = T)$.

d) Supoziția este încălcată, deoarece prin faptul că știind Sunny, aflăm Rainy și invers => Suny și Rainy nu sunt independente condițional iar BN le presupune independente.

e) $P(S=T, W=T, R=F, H=T) = \underbrace{P(\neg R \mid S, W, H)}_{1} \cdot P(H \mid S, W) =$

$= P(S, W \mid H)\, P(H) \overset{\text{Cond. indep}}{=} P(S \mid H) \cdot P(W \mid H) \cdot P(H) = 0.16$

f. Care este rata medie a erorilor produse de clasificatorul Bayes Naiv când se folosesc toate cele 3 atribute de intrare?

tributes? Does the performance of Naive Bayes improve by ... Rainy? Explain why.

Solution:

X_1	X_2	X_3	Y	$P(X_1, X_2, X_3, Y)$	$P_{NB}(X_1, X_2, X_3, Y) = P(X_3 \mid Y) \cdot P(X_1 \mid Y) \cdot P(X_2 \mid Y) \cdot P(Y)$	$Y_{NB}(X_1, X_2, X_3)$	$P_{NB}(Y \mid X_1, X_2, X_3)$
F	F	F	F	0	$0.075 \cdot 0.7 = 0.0525$	F	0.522388
F	F	F	T	**0**	$0.060 \cdot 0.8 = 0.0480$	F	0.477612
F	F	T	F	0.075	$0.075 \cdot 0.3 = 0.0225$	F	0.652174
F	F	T	T	**0.060**	$0.060 \cdot 0.2 = 0.0120$	F	0.347826
F	T	F	F	0	$0.075 \cdot 0.7 = 0.0525$	F	0.621302
F	T	F	T	**0**	$0.040 \cdot 0.8 = 0.0320$	F	0.378698
F	T	T	F	0.075	$0.075 \cdot 0.3 = 0.0225$	F	0.737705
F	T	T	T	**0.040**	$0.040 \cdot 0.2 = 0.0080$	F	0.262295
T	F	F	F	**0.175**	$0.175 \cdot 0.7 = 0.1225$	T	0.389507
T	F	F	T	0.240	$0.240 \cdot 0.8 = 0.1920$	T	0.610493
T	F	T	F	0	$0.175 \cdot 0.3 = 0.0525$	F	0.522388
T	F	T	T	**0**	$0.240 \cdot 0.2 = 0.0480$	F	0.477612
T	T	F	F	**0.175**	$0.175 \cdot 0.7 = 0.1225$	T	0.489022
T	T	F	T	0.160	$0.160 \cdot 0.8 = 0.1280$	T	0.510978
T	T	T	F	0	$0.175 \cdot 0.3 = 0.0525$	F	0.621302
T	T	T	T	**0**	$0.160 \cdot 0.2 = 0.0320$	F	0.378698

$$\text{error} \overset{def}{=} 0 + 0.175 + 0 + 0.175 + 0.040 + 0 + 0.060 + 0 =$$
$$= 0.45$$

g) $0.45 > 0.435$, performanța a scăzut, deoarece avem o dependență între atributele Rainy și Sunny, însă BN le presupune independente, lucru care este greșit.

40. **(Algoritmul Bayes Naiv: comparaţie cu alţi clasificatori)**

prelucrare de L. Ciortuz, după
* *CMU, 2004 fall, T. Mitchell, Z. Bar-Joseph, midterm, pr. 6.b*

Considerăm un clasificator Bayes Naiv care lucrează pe un set de date descrise de atributele de intrare A şi B şi atributul de ieşire Y. (Exemplu: $Y = A$ XOR B). A şi B sunt variabile aleatoare independente între ele.

a. Este posibil ca, în situaţia aceasta, vreun alt clasificator — de exemplu, ID3, regresia logistică (eventual kernel-izată) sau SVM — să lucreze mai bine decît clasificatorul Bayes Naiv?

b. Care este motivul?

a) Da este posibil

b) funcția XOR creează o relație între A și B care este non-lineară și dependentă. În cazul XOR, valoarea lui Y nu poate fi determinată decât dacă se cunosc atributele A și B. => BN face presupunerea că sunt indep. ceea ce este greșit =>

=) De aceea BN nu va performa foarte bine.

regresia logistică kernel-izată, SVM și ID3, vor performa mult mai bine, deoarece acești algoritmi pot gestiona relația non-liniară mai bine decât o face BN.

0.2.4 Estimarea parametrilor unor distribuţii probabiliste

123. **(Distribuţia Bernoulli: câteva chestiuni de bază; estimarea parametrului)**

CMU, 2018 spring, Nina Balcan, HW0, pr. A.3.1-4

Considerăm setul de date $S = \{1, 1, 0, 1, 0\}$, obţinut prin aruncarea unei monede X de cinci ori, unde 0 indică faptul că la aruncarea monedei a fost obţinută faţa cu banul, iar 1 semnifică faptul că la aruncarea monedei a fost obţinută faţa cu stema.

a. Calculaţi $\bar{x} = \frac{1}{|S|}\sum_{x_i \in S} x_i$, *media empirică* (sau, media la eşantionare; engl., the sample mean) pentru acest set de date.

b. Calculaţi $\sigma^2 = \frac{1}{|S|}\sum_{x_i \in S}(x_i - \bar{x})^2$, *varianţa empirică* (sau, varianţa la eşantionare; engl., the sample variance) pentru acest set de date.

c. Calculaţi probabilitatea de a „observa“ / obţine aceste date, presupunând că ele au fost generate prin aruncarea unei monede perfecte, adică având distribuţia de probabilitate $P(X = 1) = 0.5$ şi $P(X = 0) = 0.5$.

d. Remarcaţi faptul că probabilitatea [generării] acestui set de date ar fi putut fi mai mare dacă valoarea lui $P(X = 1)$ n-ar fi fost 0.5, ci alta. Care este valoarea care maximizează probabilitatea [generării] setului S? Justificaţi riguros.

a) $$\bar{X} = \frac{1}{|S|}\sum_{x_i \in S} x_i = \frac{1}{5}\cdot(1+1+0+1+0) = \frac{3}{5}$$

b) $$\sigma^2 = \frac{1}{|S|}\sum_{x_i \in S}(x_i - \bar{X})^2 = \frac{1}{5}\left(\left(1-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2 + \left(1-\frac{3}{5}\right)^2 + \left(0-\frac{3}{5}\right)^2\right)$$

$$= \frac{1}{5}\left(\frac{4}{25} + \frac{4}{25} + \frac{9}{25} + \frac{4}{25} + \frac{9}{25}\right) =$$

$$= \frac{1}{\cancel{5}}\cdot\frac{\overset{6}{\cancel{30}}}{25} = \frac{6}{25}$$

c)

$$P(S) = P(X=1)\cdot P(X=1)\cdot P(X=0)\cdot P(X=1)\cdot P(X=0) =$$

$$= 0.5^5 = 0.03123$$

d) Folosind estimarea Verosimilității Maxime.

Definim parametrul pe care vrem să-l calculăm: $p = P(X=1)$

prob. să pice stemă.

iar prob. să pice ban =) $1-p$.

$$P(S \mid p) = P(X=1, X=1, X=0, X=1, X=0 \mid X=1) = \quad \text{ind. cond}$$
$$= P(X=1 \mid X=1)^3 \cdot P(X=0 \mid X=1)^2 =$$
$$= p^3 \cdot (1-p)^2 =$$

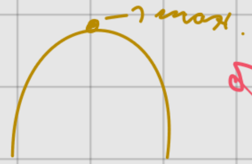
Știm fct. exprimată în fct. de $p \Rightarrow L(p) = p^3 \cdot (1-p)^2$

• folosim fct. logaritmică pt $L(p) \Rightarrow \ln L(p) = \ln p^3 + \ln(1-p)^2 =$

$$\ln L(p) = 3 \ln p + 2 \ln(1-p)$$

• derivăm pt. a afla max. $\Rightarrow \dfrac{dL}{dp} = \dfrac{d(3 \ln p + 2 \ln(1-p))}{dp} =$

$$= \frac{3}{p} - \frac{2}{(1-p)} \cdot (-1) = \underbrace{\frac{3}{p} - \frac{2}{(1-p)}}$$



→ după egalăm derivata cu 0 pt. a găsi pct. maxim $\Rightarrow$

$$\Rightarrow \frac{3}{p} - \frac{2}{(1-p)} = 0 \;/\; p(1-p) \Rightarrow 3(1-p) - 2p = 0 \Rightarrow 3 - 3p - 2p = 0 \Rightarrow$$

$\Rightarrow -5p = -3 \Rightarrow p = \dfrac{3}{5} \Rightarrow P(X=1) = \dfrac{3}{5}$ → maximul pt setul nostru de date S. | facem și a 2-a derivată $l''(p) = -\dfrac{3}{p^2} - \dfrac{2}{(1-p)^2} < 0, \forall p \Rightarrow$

$\Rightarrow$ concavă

dacă nu ne dăm seama $\Rightarrow$

$l''\left(\frac{3}{5}\right) = \dots < 0$

↳ pt. resp. MLE.

de obicei ne așteptăm să fie concavă.

?

125. (Distribuţia Bernoulli: estimarea parametrului (MLE şi MAP), într-un caz particular)

□ • ○ *CMU, 2011 fall, T. Mitchell, A. Singh, midterm exam, pr. 2*

În acest exerciţiu veţi estima probabilitatea ca la aruncarea unei monede să apară faţa *stemă* (engl., head), făcând estimări în sensul verosimilităţii maxime (MLE) şi, respectiv, în sensul probabilităţii maxime a posteriori (MAP).

Presupunem că dispunem de o monedă pentru care probabilitatea de apariţie a feţei cu stema este $p = 0.5$, adică este o monedă perfectă (engl., fair coin). Totuşi, am dori să-i calculăm un estimator, $\hat{\theta}$. În mod obişnuit, un astfel de estimator — pentru o distribuţie de tip Bernoulli — se calculează presupunând că el poate lua orice valoare din intervalul $[0, 1]$ (vedeţi problema **43.c** sau problema **124.a**). Aici, în schimb, vom impune ca $\hat{\theta}$ să ia valori într-o mulţime finită, şi anume $\{0.3, 0.6\}$.

a. Presupunem că am aruncat moneda de 3 ori şi am obţinut de 2 ori *banul* şi o dată *stema*. Calculaţi estimarea de verosimilitate maximă $\hat{\theta}_{MLE}$ a lui p în raport cu setul de valori posibile, $\{0.3, 0.6\}$.[243]

$\mapsto$ să fie stemă

a) $p = P(X=1) = 0.5$

[deoarece repetăm experimentul de 3 ori] $S = \{0, 0, 1\}$, $p \in \{0.3, 0.6\}$

$L(\theta) = 3 \cdot P(\text{Date} \mid \theta) = P(X=0, X=0, X=1 \mid \theta) =$

$= 3\, P(X=0 \mid \theta) \cdot P(X=0 \mid \theta) \cdot P(X=1 \mid \theta) =$

$= 3 \cdot \theta \cdot (1-\theta)^2$

X luăm pe rând $\theta \in \{0.3, 0.6\} \Rightarrow \theta = 0.3 \Rightarrow 3 \cdot \left(\frac{3}{10}\right) \cdot \left(1 - \frac{3}{10}\right)^2 = 3 \cdot \frac{3}{10}\left(\frac{7}{10}\right)^2 = 0.147 \cdot 3$

$= 0.441$

e invers $\left(\frac{3}{10}\right)^2 \cdot \left(1 - \frac{3}{10}\right) \ldots$

eroi.

$\Rightarrow \theta = 0.6 \Rightarrow 3 \cdot \frac{6}{10}\left(1 - \frac{6}{10}\right)^2 = 3 \cdot 0.096 = 0.288$

estimarea de verosimilitate maximă a lui p este pt $p = 0.3$.

?

b. Presupunem că folosim următoarea distribuţie de probabilitate a priori pentru valorile parametrului p:

$$P(p = 0.3) = 0.3 \text{ şi } P(p = 0.6) = 0.7.$$

Presupunem din nou că am aruncat moneda de 3 ori şi am obţinut de 2 ori *banul* şi o dată *stema*. Calculaţi estimarea de probabilitate maximă a posteriori $\hat{\theta}_{MAP}$ a lui p în raport cu mulţimea de valori posibile, $\{0.3,\ 0.6\}$, folosind această distribuţie a priori.

$P(D \mid \theta)\, P(\theta)$

$S = \{0, 0, 1\}$ $p \in \{0.3, 0.6\}$

$L(\hat{\theta}_{MAP}) = 3 \cdot P(\text{Date} \mid p) \cdot \underbrace{P(p)}_{\text{def.}} = 3 \cdot P(X=0, X=0, X=1 \mid p) \cdot P(p) =$

$= 3 \cdot P(X=0 \mid p) \cdot P(X=0 \mid p) \cdot P(X=1 \mid p) \cdot P(p)$

$3 \cdot P(X=0 \mid p) \cdot P(X=0 \mid p) \quad P(X=1 \mid p) \cdot P(p) =$

$= 3 \cdot p \cdot (1-p)^2 \cdot P(p)$

pt. $p = 0.3 \Rightarrow 3 \cdot \left(\frac{3}{10}\right)^2 \cdot \left(\frac{7}{10}\right)^{\cancel{2}} \frac{3}{10} = \cancel{\frac{27}{100} \cdot \frac{49}{100}} = \cancel{0.1323}$

$p = 0.6 \Rightarrow 3 \cdot \left(\frac{6}{10}\right)^2 \cdot \cancel{\left(\frac{6}{10}\right)^2} \cdot \frac{7}{10} = \frac{18 \cdot 16 \cdot 7}{\cancel{100 \cdot 100}} = \cancel{0.2016}$

estimarea MAP va fi pt. $p = 0.6$.

?

d. **Presupunem din nou că aruncăm moneda de un număr de ori care tinde la infinit. Calculaţi estimarea de probabilitate maximă a posteriori $\hat{\theta}_{MAP}$ a lui p în raport cu mulţimea de valori posibile, $\{0.3,\ 0.6\}$, folosind distribuţia de probabilitate a priori care a fost definită la punctul b.**

$P(X=1)$.

se aruncă moneda de n ori, unde $n \to \infty$

$\vec{\theta}_{MAP} = m \cdot P(\text{date} \mid \theta) \cdot P(\theta) = m \cdot P(X=1, X=0 \mid p) \cdot P(p) =$

ştiind că $n \to \infty$, atunci estimarea mai mare va fi cea cu priorul mai mare $\Rightarrow \vec{\theta}_{MAP} = 0.6$

126. **(Estimare în sens MLE pentru distribuţia Bernoulli şi distribuţia binomială: exemplificare)**

□ • ***CMU, 2018 spring, Nina Balcan, HW2, pr. 2***

Presupunem că „observăm“ valorile a n variabile aleatoare i.i.d. $X_1, \ldots, X_n$, toate urmând o [aceeaşi] distribuţie Bernoulli de parametru θ.[244] Cu alte cuvinte, ştim că pentru orice $i = 1, \ldots, n$, au loc egalităţile

$$P(X_i = 1) = \theta \text{ şi } P(X_i = 0) = 1 - \theta.$$

Scopul **nostru este să estimăm [în sensul verosimilităţii maxime, MLE] valoarea parametrului θ, pornind de la aceste valori „observate“, $X_1, \ldots, X_n$.**

a. **Scrieţi formula funcţiei de log-verosimilitate, $\ell(\theta)$. Această funcţie va depinde de variabilele aleatoare $X_1, \ldots, X_n$ şi de parametrul θ. Aduceţi expresia acestei funcţii la cea mai simplă formă posibilă. (Aşadar, nu vă limitaţi la a scrie doar definiţia funcţiei de log-verosimilitate.)**

Determinaţi o formulă analitică (engl., a closed form expression) pentru estimarea de verosimilitate maximă, $\hat{\theta}_{MLE}$.[245]

Folosiţi formula pe care aţi determinat-o mai sus pentru a calcula θ_{MLE} pentru următoarele 10 „observaţii“:

$$X = (0, 1, 1, 1, 1, 0, 1, 0, 1, 1).$$

a)

$L(\theta) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m \mid \theta) =$

$= \prod_{i=1}^{m} P(X_i = x_i \mid \theta) = \prod_{i=1}^{m} \theta^{x_i}(1-\theta)^{1-x_i}$ → fct. de verosimilitate

$\ell(\theta) = \ln L(\theta) = \ln\left(\prod_{i=1}^{m} \theta^{x_i}(1-\theta)^{1-x_i}\right) = \sum_{i=1}^{m}\left(x_i \ln\theta + (1-x_i)\ln(1-\theta)\right)$

fie notăm cu k nr. toate val $\sum_{i=1}^{m} x_i = k$

- nr. de val de 1 $\Rightarrow \sum_{i=1}^{m} x_i = k$
- nr. de val de 0 $\Rightarrow \sum_{i=1}^{m} (1-x_i) \Rightarrow (m-k)$

$\Rightarrow k\ln\theta + (m-k)\ln(1-\theta)$

Formula analitică pt. estimarea verosimilității maxime: - derivăm

- egalăm cu 0

derivata

$$\frac{\delta\, l(\theta)}{\delta \theta} = \frac{k}{\theta} + \frac{(n-k)\cdot 1}{(1-\theta)}(-1) = \frac{k}{\theta} - \frac{n-k}{(1-\theta)}$$

egalăm cu 0:

$$\frac{k}{\theta} - \frac{n-k}{(1-\theta)} = 0 \Rightarrow \frac{k}{\theta} = \frac{n-k}{(1-\theta)}$$

$$\Rightarrow k - \cancel{k\theta} = n\theta - \cancel{k\theta} \Rightarrow k = n\theta \Rightarrow \boxed{\theta_{MLE} = \frac{k}{n}}$$

$$\Rightarrow \theta = \frac{7}{10} = 0.7$$