

Bài tập Lý thuyết 8

Thông tin sinh viên

Câu 1. Cho bộ dữ liệu như sau

Input 1. X1 = Yes, X2 = No, X3 = Yes, X4 = No

Input 2. X1 = No, X2 = Yes, X3 = Sometimes, X4 = Yes

Input 3. X1 = Yes, X2 = No, X3 = Yes, X4 = Yes

Input 4. X1 = No, X2 = Yes, X3 = Yes, X4 = Yes

Câu 2. Bảng dữ liệu bên dưới mô tả thống kê về thu nhập của sinh viên sau khi tốt nghiệp ở các trường. Trong đó "*Thu nhập*" là thuộc tính phân lớp.

Input 1. X1 = Đại học, X2 = Lập Trình, X3 = 3..10

Input 2. X1 = Cao Đẳng, X2 = Quản lý dự án, X3 = <3

Thông tin sinh viên

- MSSV: 22850034
- Họ và tên: Cao Hoài Việt
- Email: viet.ch2612@gmail.com

Câu 1. Cho bộ dữ liệu như sau

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
Leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
Gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Dùng phương pháp Naïve Bayes (làm tròn laplace nếu cần), hãy cho biết kết quả phân lớp của các input sau

	Give Birth	Can Fly	Live in Water	Have Legs	Class
1	Yes	No	Yes	No	?
2	No	Yes	Sometimes	Yes	?
3	Yes	No	Yes	Yes	?
4	No	Yes	Yes	Yes	?

Ta định danh như sau

- X_1 = Give Birth
- X_2 = Can Fly
- X_3 = Live in Water
- X_4 = Have Legs
- Y = Class

$$P(Y = \text{mammals}|X) \sim P(X|Y = \text{mammals}) * P(Y = \text{mammals})$$

$$P(Y = \text{non-mammals}|X) \sim P(X|Y = \text{non-mammals}) * P(Y = \text{non-mammals})$$

$$P(Y = \text{mammals}) = 7/20$$

$$P(Y = \text{non-mammals}) = 13/20$$

Input 1. X1 = Yes, X2 = No, X3 = Yes, X4 = No

	X1=Yes	X2=No	X3=Yes	X4=No	P(Y X)
Y = mammals	6/7	6/7	2/7	2/7	$=6/7*6/7*2/7*2/7*7/20 \sim 0.02099$
Y = non-mammals	1/13	10/13	3/13	4/13	$=1/13*10/13*3/13*4/13*13/20 \sim 0.00273$

Vì $P(Y = \text{mammals}|X) \sim 0.02099 > P(Y = \text{non-mammals}) \sim 0.00273$ nên $Y = \text{mammals}$ với $X1=\text{Yes}, X2=\text{No}, X3=\text{Yes}, X4=\text{No}$

Input 2. X1 = No, X2 = Yes, X3 = Sometimes, X4 = Yes

	X1=No	X2=Yes	X3=Sometimes	X4=Yes	P(Y X)
Y = mammals	1/7	1/7	0/7	5/7	$=1/7*1/7*2/7*5/7 \sim 0.02099$
Y = non-mammals	12/13	3/13	4/13	9/13	$=1/13*10/13*3/13*4/13*13/20 \sim 0.00273$

Vì có biến bằng 0 nên chúng ta sẽ làm tròn (Laplace).

$$P(Y = \text{mammals}) = 8/22$$

$$P(Y = \text{non-mammals}) = 14/22$$

	X1=No	X2=Yes	X3=Sometimes	X4=Yes	P(Y X)
Y = mammals	2/9	2/9	1/10	6/9	$=2/9*2/9*1/10*6/9*8/22 \sim 0.00120$
Y = non-mammals	13/15	4/15	5/16	10/14	$=13/15*4/15*5/16*10/14*14/22 \sim 0.03283$

Vì $P(Y = \text{mammals}|X) \sim 0.00120 < P(Y = \text{non-mammals}) \sim 0.03283$ nên $Y = \text{non-mammals}$ với $X1=\text{No}, X2=\text{Yes}, X3=\text{Sometimes}, X4=\text{Yes}$

Input 3. X1 = Yes, X2 = No, X3 = Yes, X4 = Yes

	X1=Yes	X2=No	X3=Yes	X4=Yes	P(Y X)
Y = mammals	6/7	6/7	2/7	5/7	$=6/7*6/7*2/7*5/7*7/20 \sim 0.05248$
Y = non-mammals	1/13	10/13	3/13	9/13	$=1/13*10/13*3/13*9/13*13/20 \sim 0.00614$

Vì $P(Y = \text{mammals}|X) \sim 0.05248 > P(Y = \text{non-mammals}) \sim 0.00614$ nên $Y = \text{mammals}$ với $X1=\text{Yes}, X2=\text{No}, X3=\text{Yes}, X4=\text{Yes}$

Input 4. X1 = No, X2 = Yes, X3 = Yes, X4 = Yes

	X1=No	X2=Yes	X3=Yes	X4=Yes	P(Y X)
Y = mammals	1/7	1/7	2/7	5/7	$=1/7*1/7*2/7*5/7*7/20 \sim 0.00146$
Y = non-mammals	12/13	3/13	3/13	9/13	$=12/13*3/13*3/13*9/13*13/20 \sim 0.02212$

Vì $P(Y = \text{mammals}|X) \sim 0.00146 < P(Y=\text{non-mammals}) \sim 0.02212$ nên $Y = \text{non-mammals}$ với $X1=\text{No}$, $X2=\text{Yes}$, $X3=\text{Yes}$, $X4=\text{Yes}$

Câu 2. Bảng dữ liệu bên dưới mô tả thống kê về thu nhập của sinh viên sau khi tốt nghiệp ở các trường. Trong đó “*Thu nhập*” là thuộc tính phân lớp.

Mã SV	Trình độ học vấn	Nghề nghiệp	Số năm kinh nghiệm	Thu nhập
1712101	Cao đẳng	Lập trình	<3	Thấp
1412102	Cao đẳng	Lập trình	3...10	Thấp
1612103	Đại học	Lập trình	<3	Cao
1012104	Đại học	Quản lý dự án	>10	Thấp
1312105	Đại học	Quản lý dự án	3...10	Thấp
1412106	Đại học	Quản lý dự án	3...10	Cao
1012107	Đại học	Lập trình	>10	Cao
1512108	Đại học	Quản lý dự án	<3	Thấp
1012109	Cao đẳng	Lập trình	>10	Cao
1012110	Cao đẳng	Quản lý dự án	>10	Thấp

Hãy áp dụng phương pháp Naïve Bayes (sửa lỗi Laplace nếu cần) để xác định lớp cho hai mẫu dữ liệu sau:

Ta định danh như sau

- X1 = Trình độ học vấn
- X2 = Nghề nghiệp
- X3 = Số năm kinh nghiệm
- Y = Thu nhập

$$P(Y = \text{Thấp}|X) \sim P(X|Y = \text{Thấp}) * P(Y = \text{Thấp})$$

$$P(Y = \text{Cao}|X) \sim P(X|Y = \text{Cao}) * P(Y = \text{Cao})$$

$$P(Y = \text{Thấp}) = 6/10$$

$$P(Y = \text{Cao}) = 4/10$$

Input 1. X1 = Đại học, X2 = Lập Trình, X3 = 3..10

	X1=Đại học	X2=Lập trình	X3=3..10	P(Y X)
Y = Thấp	3/6	2/6	2/6	$=3/6*2/6*2/6*6/10 \sim 0.033333$
Y = Cao	3/4	3/4	1/4	$=3/4*3/4*1.4*4/10 = 0.315$

Vì $P(Y = \text{Thấp}|X) \sim 0.033333 < P(Y=\text{Cao}) = 0.315$ nên Y = Cao với X1 = Đại học, X2 = Lập Trình, X3 = 3..10

Input 2. X1 = Cao Đẳng, X2 = Quản lý dự án, X3 = <3

	X1=Cao Đẳng	X2=Quản lý dự án	X3=<3	P(Y X)
Y = Thấp	3/6	4/6	2/6	$=3/6*4/6*2/6*6/10 \sim 0.06667$
Y = Cao	1/4	1/4	1/4	$=1/4*1/4*1.4*4/10 = 0.035$

Vì $P(Y = \text{Thấp}|X) \sim 0.06667 > P(Y=\text{Cao}) = 0.035$ nên Y = Thấp với X1 = Cao Đẳng, X2 =Quản Lý Dự Án, X3= <3