SC4000/CZ4041/CE4041: Machine Learning

Solutions to L3 Tutorial Questions

Kelly KE
School of Computer Science and Engineering,
NTU, Singapore

Question 1.1

- Estimate conditional probabilities

$$P(x_i = k | y = c) = \frac{|(x_i = k) \wedge (y = c)|}{|y = c|}$$

Table 1: Data set for Question 1.

Record	A	B	C	Class
1	0	0	0	-
2	0	0	1	-
3	0	1	1	+
4	0	1	1	+
5	0	0	1	-
6	1	0	1	+
7	1	1	0	-
8	1	0	0	-
9	1	0	1	+
10	0	1	0	-

$\boxed{P(A=1\|-) = 2/6 = 0.33}$	$P(A=0\|-) = 4/6 = 0.67$
$P(B=1\|-) = 2/6 = 0.33$	$P(B=0\|-) = 4/6 = 0.67$
$P(C=1\|-) = 2/6 = 0.33$	$P(C=0\|-) = 4/6 = 0.67$
$P(A=1\|+) = 2/4 = 0.5$	$P(A=0\|+) = 2/4 = 0.5$
$P(B=1\|+) = 2/4 = 0.5$	$P(B=0\|+) = 2/4 = 0.5$
$P(C=1\|+) = 4/4 = 1$	$P(C=0\|+) = 0/4 = 0$

Naïve Bayes Classifier (Review)

Naïve Bayes Classifier: $P(\boldsymbol{x}|y=c)=\prod_{i=1}^{d} P(x_i|y=c)$

- To classify a test record $\boldsymbol{x}^*$, we need to compute the posteriors for each class c by using

$$P(y=c|\boldsymbol{x}^*)=\frac{P(y=c)\prod_{i=1}^{d} P(x_i^*|y=c)}{P(\boldsymbol{x}^*)}$$

- $P(\boldsymbol{x}^*)$ is constant for different c, it is sufficient to choose the class that maximizes the numerator term

$$P(y=c)\prod_{i=1}^{d} P(x_i^*|y=c)$$

Question 1.2

Task: Predict class label for $(A = 1, B = 1, C = 1)$ using NB

Let $P(A = 1, B = 1, C = 1) = K$

$P(+|A = 1, B = 1, C = 1)$

$$= \frac{P(A = 1, B = 1, C = 1| +) \times P(+)}{K}$$

$$= \frac{P(A = 1| +) \times P(B = 1| +) \times P(C = 1| +) \times P(+)}{K}$$

$$= \frac{0.5 \times 0.5 \times 1 \times 0.4}{K}$$

$$= \frac{0.1}{K}$$

Question 1.2 (cont.)

$$P(-|A=1,B=1,C=1)$$

$$=\frac{P(A=1,B=1,C=1|-)\times P(-)}{K}$$

$$=\frac{P(A=1|-)\times P(B=1|-)\times P(C=1|-)\times P(-)}{K}$$

$$=\frac{0.33\times 0.33\times 0.33\times 0.6}{K}$$

$$=\frac{0.0222}{K} \quad < \quad \frac{0.1}{K} \quad P(+|A=1,B=1,C=1)$$

Class label = "+"

Question 2

Home Owner	Marital Status	Taxable Income	Cheat
No	Married	120K	**?**

$m = 3$

$p = 1/3$ for all discrete features of class **Yes**

$p = 2/3$ for all discrete features of class **No**

P(HomO=Yes|No) = 3/7
P(HomO=No|No) = 4/7
P(HomO=Yes|Yes) = 0/3
P(HomO=No|Yes) = 1

P(Marital Status = Single|No) = 2/7
P(Marital Status = Divorced|No)=1/7
P(Marital Status = Married|No) = 4/7
P(Marital Status = Single|Yes) = 2/3
P(Marital Status = Divorced|Yes)=1/3
P(Marital Status = Married|Yes) = 0/3

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(Class = No) = 7/10
P(Class = Yes) = 3/10

P(HomO=Yes|No) = ?
P(HomO=No|No) = ?
P(HomO=Yes|Yes) = ?
P(HomO=No|Yes) = ?

P(Marital Status = Single|No) = ?
P(Marital Status = Divorced|No) = ?
P(Marital Status = Married|No) = ?
P(Marital Status = Single|Yes) = ?
P(Marital Status = Divorced|Yes)=?
P(Marital Status = Married|Yes) = ?

M-estimate

$$P(x_i = k | y = c) = \frac{|(x_i = k) \wedge (y = c)| + m \times p}{|y = c| + m}$$

$$P(\boldsymbol{x}^* | \text{Class} = \text{No}) = ? \quad P(\boldsymbol{x}^* | \text{Class} = \text{Yes}) = ?$$

Tutorial

M-estimate

- A more general estimation:

User-specified parameters

Original: $P(x_i = k|y = c) = \frac{|(x_i = k) \wedge (y = c)|}{|y = c|}$

M-estimate: $P(x_i = k|y = c) = \frac{|(x_i = k) \wedge (y = c)| + \boxed{m} \times \boxed{p}}{|y = c| + m}$

Prior information of $P(x_i = k|y = c)$ if available

For example, based on domain knowledge, you have the prior information:

Domain knowledge, not learned from data

$\tilde{P}(\text{Single}|\text{Yes}) = \frac{1}{2}$ $\tilde{P}(\text{Divorced}|\text{Yes}) = \frac{1}{3}$ $\tilde{P}(\text{Married}|\text{Yes}) = \frac{1}{6}$

In practice, prior information is not always available. Just consider m and p as hyper-parameters

Note: Laplace estimate is a special case of M-estimate when $m = n_i$ and $p = \frac{1}{n_i}$, where n_i is the number of distinct values of x_i

Home Owner	Marital Status	Taxable Income	Cheat
No	Married	120K	?

$m = 3$

$p = 1/3$ for all discrete features of class **Yes**

$p = 2/3$ for all discrete features of class **No**

P(HomO=Yes|No) = (3+2)/(7+3) = 5/10
P(HomO=No|No) = (4+2)/(7+3) = 6/10
P(HomO=Yes|Yes) = (0+1)/(3+3) = 1/6
P(HomO=No|Yes) = (3+1)/(3+3) = 4/6

P(Status=Single|No) = (2+2)/(7+3) = 4/10
P(Status=Divorced|No)= (1+2)/(7+3) = 3/10
P(Status=Married|No) = (4+2)/(7+3) = 6/10
P(Status=Single|Yes) = (2+1)/(3+3) = 3/6
P(Status=Divorced|Yes) = (1+1)/(3+3) = 2/6
P(Status=Married|Yes) = (0+1)/(3+3) = 1/6

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(Class = No) = 7/10
P(Class = Yes) = 3/10

$$\frac{|(x_i = k) \wedge (y = c)| + m \times p}{|y = c| + m}$$

$$\sum_k P(x_i = k | y = c) = 1$$

$$\hat{P}(\mathrm{HomO} = \mathrm{Yes}|\mathrm{No}) = \frac{P(\mathrm{HomO} = \mathrm{Yes}|\mathrm{No})}{P(\mathrm{HomO} = \mathrm{Yes}|\mathrm{No}) + P(\mathrm{HomO} = \mathrm{No}|\mathrm{No})} = \frac{\frac{5}{10}}{\frac{5}{10} + \frac{6}{10}} = \frac{5}{11}$$

$$\hat{P}(\mathrm{HomO} = \mathrm{No}|\mathrm{No}) = \frac{P(\mathrm{HomO} = \mathrm{No}|\mathrm{No})}{P(\mathrm{HomO} = \mathrm{Yes}|\mathrm{No}) + P(\mathrm{HomO} = \mathrm{No}|\mathrm{No})} = \frac{\frac{6}{10}}{\frac{5}{10} + \frac{6}{10}} = \frac{6}{11}$$

Home Owner	Marital Status	Taxable Income	Cheat
No	Married	120K	?

$m = 3$

$p = 1/3$ for all discrete features of class **Yes**

$p = 2/3$ for all discrete features of class **No**

P(HomO=Yes|No) = (3+2)/(7+3) = 5/10
P(HomO=No|No) = (4+2)/(7+3) = 6/10
P(HomO=Yes|Yes) = (0+1)/(3+3) = 1/6
P(HomO=No|Yes) = (3+1)/(3+3) = 4/6

P(Status=Single|No) = (2+2)/(7+3) = 4/10
P(Status=Divorced|No)= (1+2)/(7+3) = 3/10
P(Status=Married|No) = (4+2)/(7+3) = 6/10
P(Status=Single|Yes) = (2+1)/(3+3) = 3/6
P(Status=Divorced|Yes) = (1+1)/(3+3) = 2/6
P(Status=Married|Yes) = (0+1)/(3+3) = 1/6

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(Class = No) = 7/10
P(Class = Yes) = 3/10

P(HomO=Yes|No) = (5/10)/(5/10+6/10) = 5/11
P(HomO=No|No) = (6/10)/(5/10+6/10) = 6/11
P(HomO=Yes|Yes) = (1/6)/(1/6+4/6) = 1/5
P(HomO=No|Yes) = (4/6)/(1/6+4/6) = 4/5

P(Status=Single|No) = (4/10)/(4/10+3/10+6/10) = 4/13
P(Status=Divorced|No)= (3/10)/(4/10+3/10+6/10) = 3/13
P(Status=Married|No) = (6/10)/(4/10+3/10+6/10) = 6/13
P(Status=Single|Yes) = (2+1)/(3+3) = 3/6
P(Status=Divorced|Yes) = (1+1)/(3+3) = 2/6
P(Status=Married|Yes) = (0+1)/(3+3) = 1/6

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(Class = No) = 7/10
P(Class = Yes) = 3/10

Home Owner	Marital Status	Taxable Income	Cheat
No	Married	120K	?

$m = 3$

$p = 1/3$ for all discrete features of class **Yes**

$p = 2/3$ for all discrete features of class **No**

P(HomO=Yes|No) = 5/11
P(HomO=No|No) = 6/11
P(HomO=Yes|Yes) = 1/5
P(HomO=No|Yes) = 4/5

P(Status=Single|No) = 4/13
P(Status=Divorced|No) = 3/13
P(Status=Married|No) = 6/13
P(Status=Single|Yes) = 3/6
P(Status=Divorced|Yes) = 2/6
P(Status=Married|Yes) = 1/6

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

P(Class = No) = 7/10
P(Class = Yes) = 3/10

$$\begin{aligned}P(\mathbf{X}|Class=No) &= P(HomO=No \mid Class=No) \\ &\times P(Status=Married \mid Class=No) \\ &\times P(Income=120K \mid Class=No) \\ &= 6/11 \times 6/13 \times 0.0072 = 0.0018\end{aligned}$$

$$\begin{aligned}P(\mathbf{X}|Class=Yes) &= P(HomO=No \mid Class=Yes) \\ &\times P(Status=Married \mid Class=Yes) \\ &\times P(Income=120K \mid Class=Yes) \\ &= 4/5 \times 1/6 \times 1.2 \times 10^{-9} = 1.6 \times 10^{-10}\end{aligned}$$

$P(\mathbf{X}|No) \times P(No) > P(\mathbf{X}|Yes) \times P(Yes)$

Thus, $P(No|\mathbf{X}) > P(Yes|\mathbf{X})$

Cheat = No

Thank you!