School of Data and Computer Science, Sun Yat-sen University

T03 Planning and Uncertainty

17341015 陈鸿峥

2019 年 12 月 17 日

问题 1. Consider the following data. The DECISION-TREE-LEARNING algorithm will first select the attribute **Length** to split on. Finish building the decision tree, and show the computations.

Example	**Author**	**Thread**	**Length**	**Where Read**	**User Action (output)**
e1	known	new	long	home	skips
e2	unknown	new	short	work	reads
e3	unknown	follow Up	long	work	skips
e4	known	follow Up	long	home	skips
e5	known	new	short	home	reads
e6	known	follow Up	long	work	skips
e7	unknown	follow Up	short	work	skips
e8	unknown	new	short	work	reads
e9	known	follow Up	long	home	skips
e10	known	new	long	work	skips
e11	unknown	follow Up	short	home	skips
e12	known	new	long	work	skips
e13	known	follow Up	short	home	reads
e14	known	new	short	work	reads
e15	known	new	short	home	reads
e16	known	follow Up	short	work	reads
e17	known	new	short	home	reads
e18	unknown	new	short	work	reads

解答. 按照Length分裂得到如下决策树，其中下划线标识的为reads的标签

$$\boxed{\text{Length}}$$

long: 1,3,4,6,9,10,12　　short: $\underline{2},\underline{5},7,\underline{8},11,\underline{13},\underline{14},\underline{15},\underline{16},\underline{17},\underline{18}$

由于左子树全为skips，故不用继续分裂；考虑右子树，分别对剩下的三个属性进行分裂，得到如下决策树

$\boxed{\text{Author}}$ known: $\underline{5},\underline{13},\underline{14},\underline{15},\underline{16},\underline{17}$　　unknown: $\underline{2},7,\underline{8},11,\underline{18}$

$\boxed{\text{Thread}}$ new: $\underline{2},\underline{5},\underline{8},\underline{14},\underline{15},\underline{17},\underline{18}$　　follow Up: $7,11,\underline{13},\underline{16}$

$\boxed{\text{Where Read}}$ home: $\underline{5},11,\underline{13},\underline{15},\underline{17}$　　work: $\underline{2},7,\underline{8},\underline{14},\underline{16},\underline{18}$

分别讨论其信息增益，先计算根节点的信息熵

$$Ent(D) = -\left(\frac{2}{11}\log_2\frac{2}{11} + \frac{9}{11}\log_2\frac{9}{11}\right) = 0.684038$$

然后计算每种划分的信息熵与信息增益

$$
\begin{aligned}
Gain(D, \text{Author}) &= Ent(D) - \sum_{v=1}^{2} Ent(D^v) \\
&= Ent(D) - \left(-\frac{6}{11}\left(\frac{6}{6}\log_2\frac{6}{6} + \frac{0}{6}\log_2\frac{0}{6}\right) + -\frac{5}{11}\left(\frac{3}{5}\log_2\frac{3}{5} + \frac{2}{5}\log_2\frac{2}{5}\right)\right) \\
&= 0.242697 \\
Gain(D, \text{Thread}) &= Ent(D) - \sum_{v=1}^{2} Ent(D^v) \\
&= Ent(D) - \left(-\frac{7}{11}\left(\frac{7}{7}\log_2\frac{7}{7} + \frac{0}{7}\log_2\frac{0}{7}\right) + -\frac{4}{11}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)\right) \\
&= 0.320402 \\
Gain(D, \text{Where Read}) &= Ent(D) - \sum_{v=1}^{2} Ent(D^v) \\
&= Ent(D) - \left(-\frac{5}{11}\left(\frac{4}{5}\log_2\frac{4}{5} + \frac{1}{5}\log_2\frac{1}{5}\right) + -\frac{6}{11}\left(\frac{1}{6}\log_2\frac{1}{6} + \frac{5}{6}\log_2\frac{5}{6}\right)\right) \\
&= 0.001331
\end{aligned}
$$

因此选择Thread进行进一步划分，左子树全为reads，不需划分；只需划分右子树，选择剩下两个属性可以得到

$\boxed{\text{Author}}$: skips / → 7,11; \ reads → <u>13</u>,<u>16</u>

$\boxed{\text{Where Read}}$: home / → 11,<u>13</u>; \ work → 7,<u>16</u>

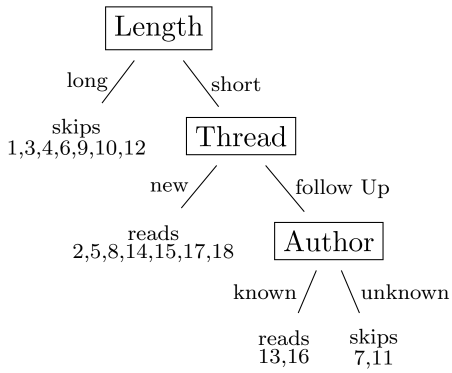
计算根节点的信息熵

$$
Ent(D) = -\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) = 1
$$

然后计算每种划分的信息熵与信息增益

$$
\begin{aligned}
Gain(D, \text{Author}) &= Ent(D) - \sum_{v=1}^{2} Ent(D^v) \\
&= Ent(D) - \left(-\frac{1}{2}\left(\frac{2}{2}\log_2\frac{2}{2} + \frac{0}{2}\log_2\frac{0}{2}\right) + -\frac{1}{2}\left(\frac{0}{0}\log_2\frac{0}{0} + \frac{2}{0}\log_2\frac{2}{0}\right)\right) \\
&= 1 \\
Gain(D, \text{Where Read}) &= Ent(D) - \sum_{v=1}^{2} Ent(D^v) \\
&= Ent(D) - \left(-\frac{1}{2}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right) + -\frac{1}{2}\left(\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right)\right) \\
&= 0
\end{aligned}
$$

因而选择Author作为划分属性。又划分之后的两个子树内元素均属于同一类别，无法继续划分，决策树算法终止。最终得到决策树如下



问题 2. Consider the candy example from the lecture. Assume that the prior distribution over $h1, \dots, h5$ is given by $\langle 0.1, 0.2, 0.4, 0.2, 0.1 \rangle$. Suppose that the first 5 candies taste lime, cherry, cherry, lime, and lime. Make predictions for the 6th candy using Bayesian, MAP and ML learning, respectively. Show the computations done to make the predictions.

解答. 题目中的假设H为

h_1	100% cherry
h_2	75% cherry + 25% lime
h_3	50% cherry + 50% lime
h_4	25% cherry + 75% lime
h_5	100% lime

给定先验$P(h_i)$和似然$P(d \mid H)$

hypothesis	h_1	h_2	h_3	h_4	h_5
$P(lime \mid h_i)$	0	0.25	0.5	0.75	1
$P(cherry \mid h_i)$	1	0.75	0.5	0.25	0
$P(h_i)$	0.1	0.2	0.4	0.2	0.1
$P(d \mid h_i)$	$(1)^2(0)^3$ $= 0$	$(0.75)^2(0.25)^3$ $= 9/1024$	$(0.5)^2(0.5)^3$ $= 1/32$	$(0.25)^2(0.75)^3$ $= 27/1024$	$(0)^2(1)^3$ $= 0$

和证据集

$$d = \langle lime, cherry, cherry, lime, lime \rangle$$

并做独立性假设$P(d \mid h) = \prod_j P(d_j \mid h)$。

(a) 贝叶斯学习：由全概率公式

$$\begin{aligned} P(d) &= \sum_i P(d \mid h_i) P(h_i) \\ &= 9/1024 \cdot 0.2 + 1/32 \cdot 0.4 + 27/1024 \cdot 0.2 \\ &= 0.01953125 \end{aligned}$$

进而

$$\begin{aligned}
P(lime \mid d) &= \sum_i P(lime \mid h_i) P(h_i \mid d) \\
&= \frac{1}{P(d)} \sum_i P(lime \mid h_i) P(d \mid h_i) P(h_i) \\
&= 0.01064453125 \\
P(cherry \mid d) &= \sum_i P(cherry \mid h_i) P(h_i \mid d) \\
&= \frac{1}{P(d)} \sum_i P(cherry \mid h_i) P(d \mid h_i) P(h_i) \\
&= 0.00888671875
\end{aligned}$$

因$P(lime \mid d) > P(cherry \mid d)$，故判为lime。

(b) 极大后验(MAP)

$$h_{MAP} = \arg\max_{h_i} P(h_i \mid d) = \arg\max_{h_i} P(h_i) P(d \mid h_i)$$

可求得

$$\begin{aligned}
P(h_1) P(d \mid h_1) &= 0 \\
P(h_2) P(d \mid h_2) &= 0.0017578125 \\
P(h_3) P(d \mid h_3) &= 0.0125 \\
P(h_4) P(d \mid h_4) &= 0.0052734375 \\
P(h_5) P(d \mid h_5) &= 0
\end{aligned}$$

故$h_{MAP} = h_3$，

$$P(lime \mid d) \approx P(lime \mid h_{MAP}) = 0.5$$
$$P(cherry \mid d) \approx P(cherry \mid h_{MAP}) = 0.5$$

(c) 极大似然(ML)

$$h_{ML} = \arg\max_{h_i} P(d \mid h_i) = h_3$$

进而

$$P(lime \mid d) \approx P(lime \mid h_{ML}) = 0.5$$
$$P(cherry \mid d) \approx P(cherry \mid h_{ML}) = 0.5$$

问题 3. Consider the Boolean function $E = (A \text{ XOR } B) \text{ AND } (C \text{ XOR } D)$. Construct its truth table, and then remove the line for the input $A = 1, B = 1, C = 1, D = 1$. Use Naive Bayes classification to make prediction for this input. Show the computations.

解答. 真值表如下：

A	B	C	D	E
1	1	1	1	0

1	1	1	0	0
1	1	0	1	0
1	1	0	0	0
1	0	1	1	0
1	0	1	0	1
1	0	0	1	1
1	0	0	0	0
0	1	1	1	0
0	1	1	0	1
0	1	0	1	1
0	1	0	0	0
0	0	1	1	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	0

由朴素贝叶斯

$$
\begin{aligned}
P(E \mid A, B, C, D) &= \frac{P(A, B, C, D \mid E)P(E)}{P(A, B, C, D)} \\
&= \frac{P(A \mid E)P(B \mid E)P(C \mid E)P(D \mid E)P(E)}{P(A, B, C, D)} \\
&= \frac{1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 1/4}{1/16} \\
&= \frac{1}{4} \\
P(\neg E \mid A, B, C, D) &= \frac{P(A, B, C, D \mid \neg E)P(\neg E)}{P(A, B, C, D)} \\
&= \frac{P(A \mid \neg E)P(B \mid \neg E)P(C \mid \neg E)P(D \mid \neg E)P(\neg E)}{P(A, B, C, D)} \\
&= \frac{1/2 \cdot 1/2 \cdot 1/2 \cdot 1/2 \cdot 3/4}{1/16} \\
&= \frac{3}{4}
\end{aligned}
$$

因为$P(\neg E \mid A, B, C, D) > P(E \mid A, B, C, D)$，故判别为$E = 0$。

问题 4. Construct a neural network that computes the XOR function of two inputs.

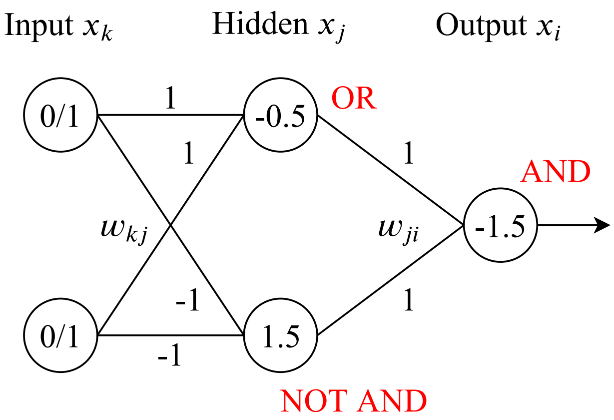
解答. XOR的真值表如下

A	B	C
0	0	0
0	1	1
1	0	1
1	1	0

注意到

$$A \oplus B = (A \vee B) \wedge (\neg(A \wedge B))$$

故可以通过2-2-1的三层神经网络实现，其中输入层对应着A, B，隐含层的两个神经元对应着OR和NOT AND的操作，输出层操作则为AND。进而可构造网络及对应权值如下图所示，其中隐含层及输出层神经元内的数值为偏置(bias)。



并且令隐含层和输出层的激活函数都为

$$\sigma(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

则前向传播规则为

$$f(x_1, x_2) = x_i = \sigma\left(\sum_j w_{ji}\sigma\left(\sum_k w_{kj}x_k + b_j\right) + b_i\right)$$

进而

$$\begin{aligned}
f(0,0) &= \sigma(\sigma(0*1+0*1-0.5)*1+\sigma(0*(-1)+0*(-1)+1.5)*1-1.5) = \sigma(0+1-1.5) = 0 \\
f(0,1) &= \sigma(\sigma(0*1+1*1-0.5)*1+\sigma(0*(-1)+1*(-1)+1.5)*1-1.5) = \sigma(1+1-1.5) = 1 \\
f(1,0) &= \sigma(\sigma(1*1+0*1-0.5)*1+\sigma(1*(-1)+0*(-1)+1.5)*1-1.5) = \sigma(1+1-1.5) = 1 \\
f(1,1) &= \sigma(\sigma(1*1+1*1-0.5)*1+\sigma(1*(-1)+1*(-1)+1.5)*1-1.5) = \sigma(1+0-1.5) = 0
\end{aligned}$$

即为所求的XOR网络。