人工智能理论作业 4

数据科学与计算机学院

18340082 邝金熙

1. 决策树

首先计算整个数据集的信息熵：一共有 18 个样本，分类为 skips 的有 9 个，为 reads 的有 9 个，概率均为 0.5. 故 $H_0 = -(\frac{1}{2}log_2\frac{1}{2} + \frac{1}{2}log_2\frac{1}{2}) = 1$.

若根据作者是否有名进行分类，则可计数得 18 个样本中有 12 个作者有名，6 个作者不出名；12 个作者有名的样本中 3 个为 skips,6 个为 reads；6 个作者有名的样本中 3 个为 skips,3 个为 reads；从而可以计算对作者是否有名进行分裂之后的信息增益：

$$\begin{aligned}
Gain_{0A} &= H_0 - H_{0A} \\
&= H_0 - (\frac{12}{18}(-(\frac{6}{12}log_2\frac{6}{12} + \frac{6}{12}log_2\frac{6}{12})) + \frac{6}{18}(-(\frac{3}{6}log_2\frac{3}{6} + \frac{3}{6}log_2\frac{3}{6}))) \\
&= 0
\end{aligned}$$

同理根据该书是新作还是续作、该书的篇幅长短、该书的阅读场所是在家还是工作地，依次计算其信息增益可以得到：

$$\begin{aligned}
Gain_{0B} &= H_0 - H_{0B} \\
&= H_0 - (\frac{10}{18}(-(\frac{3}{10}log_2\frac{3}{10} + \frac{7}{10}log_2\frac{7}{10})) + \frac{8}{18}(-(\frac{6}{8}log_2\frac{6}{8} + \frac{2}{8}log_2\frac{2}{8}))) \\
&= 0.1498258895566673 \\
Gain_{0C} &= H_0 - H_{0C} \\
&= H_0 - (\frac{7}{18}(-(\frac{7}{7}log_2\frac{7}{7} + \frac{0}{7}log_2\frac{0}{7})) + \frac{11}{18}(-(\frac{2}{11}log_2\frac{2}{11} + \frac{9}{11}log_2\frac{9}{11}))) \\
&= 0.581968199340573 \\
Gain_{0D} &= H_0 - H_{0D} \\
&= H_0 - (\frac{8}{18}(-(\frac{4}{8}log_2\frac{4}{8} + \frac{4}{8}log_2\frac{4}{8})) + \frac{10}{18}(-(\frac{5}{10}log_2\frac{5}{10} + \frac{5}{10}log_2\frac{5}{10}))) \\
&= 0
\end{aligned}$$

注意到其中 $1log_2 1 = 0, 0log_2 0 = 0$，第二个等式是因为 $\frac{1}{x}$ 是 $log_2 x$ 的高阶无穷小。

比较四个信息增益可以发现按篇幅长短分裂得到的增益最大，分裂出两个子节点：篇幅长的节点包含 7 个样本，全部为 skips 类，不需要再分裂；篇幅短的节点包含 9 个样本，其中 2 个 skips,9 个 reads，因此需要再次分裂。重复上述步骤计算信息增益：

$$H_1 = -(\frac{2}{11}log_2\frac{2}{11} + \frac{9}{11}log_2\frac{9}{11}) = 0.6840384356390418$$

$$\begin{aligned}
Gain_{1A} &= H_1 - H_{1A} \\
&= H_1 - (\frac{6}{11}(-(\frac{0}{6}log_2\frac{0}{6} + \frac{6}{6}log_2\frac{6}{6})) + \frac{5}{11}(-(\frac{2}{5}log_2\frac{2}{5} + \frac{3}{5}log_2\frac{3}{5}))) \\
&= 0.24268559765261632 \\
Gain_{1B} &= H_1 - H_{1B} \\
&= H_1 - (\frac{7}{11}(-(\frac{0}{7}log_2\frac{0}{7} + \frac{7}{7}log_2\frac{7}{7})) + \frac{4}{11}(-(\frac{2}{4}log_2\frac{2}{4} + \frac{2}{4}log_2\frac{2}{4}))) \\
&= 0.3203884701990212 \\
Gain_{1D} &= H_0 - H_{1D} \\
&= H_1 - (\frac{5}{11}(-(\frac{1}{5}log_2\frac{1}{5} + \frac{4}{5}log_2\frac{4}{5})) + \frac{6}{11}(-(\frac{1}{6}log_2\frac{1}{6} + \frac{5}{6}log_2\frac{5}{6}))) \\
&= 0.0013316170638657532
\end{aligned}$$

比较三个信息增益可以发现按新作还是续作分裂得到的增益最大，分裂出两个子节点：新作的节点包含 7 个样本，全部为 reads 类，不需要再分裂；续作的节点包含 4 个样本，其中 2 个 skips,2 个 reads，因此需要再次分裂。重复上述步骤计算信息增益：

$$H_2 = -(\frac{2}{4}log_2\frac{2}{4} + \frac{2}{4}log_2\frac{2}{4}) = 1$$

$$\begin{aligned}
Gain_{2A} &= H_2 - H_{2A} \\
&= H_2 - (\frac{2}{4}(-(\frac{0}{2}log_2\frac{0}{2} + \frac{2}{2}log_2\frac{2}{2})) + \frac{2}{4}(-(\frac{2}{2}log_2\frac{2}{2} + \frac{0}{2}log_2\frac{0}{2}))) \\
&= 1 \\
Gain_{2D} &= H_2 - H_{2D} \\
&= H_2 - (\frac{2}{4}(-(\frac{1}{2}log_2\frac{1}{2} + \frac{1}{2}log_2\frac{1}{2})) + \frac{2}{4}(-(\frac{1}{2}log_2\frac{1}{2} + \frac{1}{2}log_2\frac{1}{2}))) \\
&= 0
\end{aligned}$$

比较两个信息增益可以发现按作者是否出名得到的增益最大，分裂出两个子节点：作者出名的节点包含 2 个样本，全部为 reads 类，不需要再分裂；作者不出名的节点包含 2 个样本，全部为 skips 类，不需要再分裂；最终得到的决策树如图 1 所示，边上的文字为判断依据，圈内的文字为所预测的类：

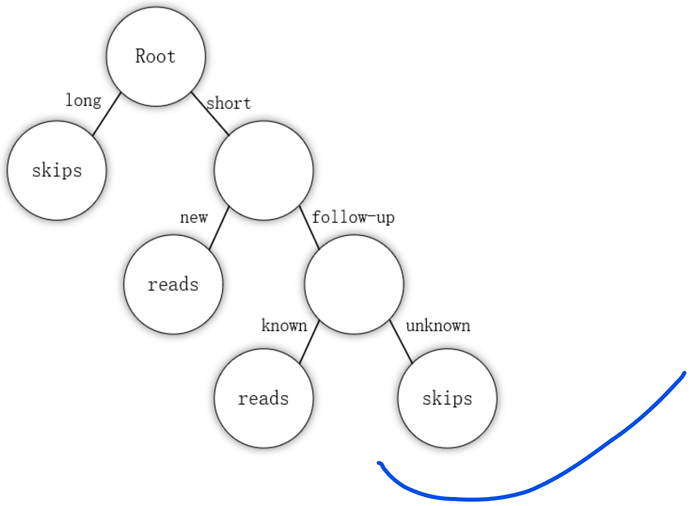

图 1: 决策树

2. 概率论推理

Notation: 本题中 h_i 表示第 i 个假设，d 表示已经出现的事实，X 表示预测的类别。

(a) Bayesian

根据贝叶斯推理，我们有：

$$P(h_i|d) = \frac{P(h_i,d)}{P(d)} = \frac{P(d|h_i)*P(h_i)}{P(d)}$$

$$\begin{aligned} P(X|d) &= \sum_i P(X|h_i)P(h_i|d) \\ &= \sum_i P(X|h_i)\frac{P(d|h_i)*P(h_i)}{P(d)} \\ &= \frac{\sum_i P(X|h_i)P(d|h_i)P(h_i))}{\sum_i P(d|h_i)P(h_i))} \end{aligned}$$

其中 $P(h_i)$ 为每种假设的先验概率，根据题设依次为 0.1,0.2,0.4,0.2,0.1；$P(X|h_i)$ 为每种假设下猜测为樱桃味或青柠味的概率，如 $P(cherry|h_1) = 100\%$；$P(d|h_i)$ 为在该假设下出现 d 的概率，本题中 d=[lime, cherry, cherry, lime, lime]，例 $P(d|h_1) = 1^2 * 0^3 = 0$，$P(d|h_2) = 0.75^2 * 0.25^3 = 0.008789$。

根据上述规则代入本题数据得

$$P(cherry|d) = 0.455$$
$$P(lime|d) = 0.545$$

即在贝叶斯学习中下一颗为青柠味的可能性略大，预测下一颗出现的是青柠味的糖果。

(b) MAP

在最大后验方法中，比较各种假设在考虑先验概率后的，在当前事实下出现的可能性，取可能性最大的一种作为当前假设，再根据该假设预测。即

$$h_{MAP} = argmax_h P(h_i|d) = argmax_h \frac{P(d|h_i)P(h_i)}{P(d)}$$

其中对于所有假设 P(d) 均相同，故只需要比较分子部分。代入数据得：

$$P(d|h_1)P(h_1) = 0$$
$$P(d|h_2)P(h_2) = 0.0017578125$$
$$P(d|h_3)P(h_3) = 0.0125$$
$$P(d|h_4)P(h_4) = 0.0052734375$$
$$P(d|h_5)P(h_5) = 0$$

从而有 $h_{MAP} = h_3$，在假设 h_3 中，出现樱桃味和青柠味的概率均为 50%，故预测两种情况的出现概率相等。

(c) ML

在最大似然方法中，比较当前事实在各种假设下出现的可能性，取可能性最大的一种作为当

前假设，再根据该假设预测。即

$$h_{ML} = argmax_h P(d|h_i)$$

代入数据得：

$$P(d|h_1) = 0$$
$$P(d|h_2) = 0.0087890625$$
$$P(d|h_3) = 0.03125$$
$$P(d|h_4) = 0.0263671875$$
$$P(d|h_5) = 0$$

从而有 $h_{ML} = h_3$，在假设 h_3 中，出现樱桃味和青柠味的概率均为 50%，故预测两种情况的出现概率相等。

3. 朴素贝叶斯分类

首先构建真值表如表 1 所示。

不妨记事件 $A = 1$ 为 a，$A = 0$ 为 $\neg a$，则：

$$\begin{aligned}
P(e|abcd) &= \frac{P(abcde)}{P(abcd)} \\
&= \frac{P(a|e)P(b|e)P(c|e)P(d|e)P(e)}{P(abcd)} \\
P(\neg e|abcd) &= \frac{P(abcd\neg e)}{P(abcd)} \\
&= \frac{P(a|\neg e)P(b|\neg e)P(c|\neg e)P(d|\neg e)P(\neg e)}{P(abcd)}
\end{aligned}$$

由于两式分母相同，故只需要比较分母。在已有的 15 个样本中 E=1 的有 4 个，E=0 的有 11 个，故 $P(e) = \frac{4}{15}, P(\neg e) = \frac{11}{15}$；在 E=1 的 4 个样本中，A=1 的样本有 2 个，故 $P(a|e) = \frac{2}{4}$；在 E=0 的 11 个样本中，A=1 的样本有 5 个，故 $P(a|\neg e) = \frac{5}{11}$；同理可数得 $P(b|e) = P(c|e) = P(d|e) = \frac{2}{4}$，

A,B,C,D,E				
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0

表 1: 真值表

$P(b|\neg e) = P(c|\neg e) = P(d|\neg e) = \frac{5}{11}$，代入上述数据得：

$$P(e|abcd) = \frac{0.016666666666666666}{P(abcd)}$$
$$P(\neg e|abcd) = \frac{0.03130478337089907}{P(abcd)}$$

从而有 $P(e|abcd) < P(\neg e|abcd)$，即预测当输入为 A=B=C=D=1 时，输出 E=0。该预测结果与真实结果 $1 \oplus 1 \wedge 1 \oplus 1 = 0$ 吻合。

4. 异或神经网络

根据布尔代数知识，我们有 $A \oplus B = A\overline{B} + \overline{A}B$，从而我们可以构建类似 ML 课件 3 第 24 页的主析取范式网络，即隐藏层神经元计算合取，输出层神经元计算析取，神经元中的激活函数采用 0-1 示性函数，即输入大于 0 输出 1，否则输出 0，如图 2 所示：

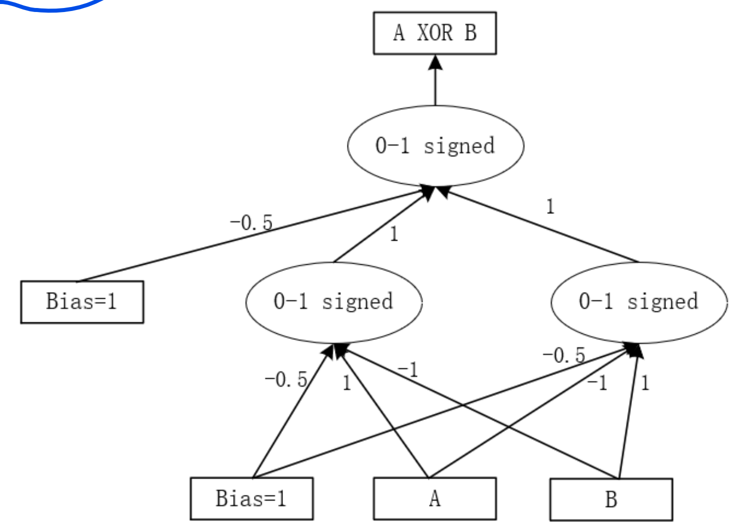

图 2: 异或神经网络

下列表验证该神经网络的正确性，每一列分别表示 A，B，隐藏层左侧神经元输入，隐藏层左侧神经元输出，隐藏层右侧神经元输入，隐藏层右侧神经元输出，输出层神经元输入，输出层神经元输出：

A	B	In1	Out1	In2	Out2	In3	Out3
0	0	-0.5	0	-0.5	0	-0.5	0
0	1	-1.5	0	0.5	1	0.5	1
1	0	0.5	1	-1.5	0	0.5	1
1	1	-0.5	0	-0.5	0	-0.5	0

表 2: 异或神经网络验证表

可以看到该神经网络的真值表与异或运算的真值表相同，符合题目要求。

5. 计算损失对权重的偏导

首先推导任一损失对任一权重的偏导，此处 g() 为激活函数，a_k 为输出层第 k 个神经元的输出，y_k 表示输出层第 k 个神经元的真实结果，$Loss_k$ 表示第 k 个输出层神经元的损失，损失函数使用平方损失函数，ω_{ij} 表示起点为该层第 i 个神经元，终点为该层第 j 个神经元的边的权重，与题目中的表示方法不同。使用 i_0, j_0 是为了避免与求和时的循环变量 i,j 混淆，造成像课件中的理解困难。对于只有一个隐藏层的网络，i 表示输入层计数，j 表示隐藏层计数，k 表示输出层计数：

$$
\begin{aligned}
\frac{\partial Loss_k}{\partial \omega_{i_0 j_0}} &= \frac{\partial ||y_k - a_k||^2}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)\frac{\partial g(\sum_j \omega_{jk} a_j)}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\frac{\partial \sum_j \omega_{jk} a_j}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\frac{\partial \omega_{j_0 k} a_{j_0}}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\omega_{j_0 k}\frac{\partial a_{j_0}}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\omega_{j_0 k}\frac{\partial g(\sum_i \omega_{i j_0} a_i)}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\omega_{j_0 k} g'(\sum_i \omega_{i j_0} a_{i_0})\frac{\partial \omega_{i_0 j_0} a_i}{\partial \omega_{i_0 j_0}} \\
&= -2(y_k - a_k)g'(\sum_j \omega_{jk} a_j)\omega_{j_0 k} g'(\sum_i \omega_{i j_0} a_i) a_{i_0}
\end{aligned}
$$

(a) $\partial Loss_{o1}/\partial \omega_1$

要计算的 $Loss_{o1}$ 相当于上述推导中的 $Loss_1$，ω_1 相当于 $\omega_{i=1,j=1}$。

首先计算得到 $a_{k=1} = 0.751365$，从课件题目图片中可以得到 $y_1 = 0.01$，输出层第 1 个神经元的输入 $\sum_j \omega_{j1} a_j = 1.105906$，对 sigmoid 函数的导数为 $g'(\sum_j \omega_{j1} a_j) = g(\sum_j \omega_{j1} a_j) * (1 - g(\sum_j \omega_{j1} a_j)) = 0.186816$，$\omega_{j=1,k=1} = \omega_5 = 0.4$，隐藏层第 1 个神经元的输入 $\sum_i \omega_{i1} a_i = 0.3775$，对 sigmoid 函数的导数为 $g'(\sum_i \omega_{i1} a_i) = g(\sum_i \omega_{i1} a_i) * (1 - g(\sum_i \omega_{i1} a_i)) = 0.2413$，$a_1 = 0.05$，因此：

$$
\frac{\partial Loss_{o1}}{\omega_1} = -2(y_1 - a_1)g'(\sum_j \omega_{j1} a_j)\omega_{j=1,k=1} g'(\sum_i \omega_{i1} a_i) a_1 = 0.0013367920422887463
$$

(b) $\partial Loss_{o2}/\partial\omega_4$

要计算的 $Loss_{o2}$ 相当于上述推导中的 $Loss_2$，ω_4 相当于 $\omega_{i=2,j=2}$。

首先计算得到 $a_{k=2}=0.755531$, 从课件题目图片中可以得到 $y_1=0.99$, 输出层第 2 个神经元的输入 $\sum_j \omega_{j2}a_j=0.985721$, 对 tanh 函数的导数为 $g'(\sum_j \omega_{j2}a_j)=1-g(\sum_j \omega_{j2}a_j)^2=0.429171$, $\omega_{j=2,k=2}=\omega_8=0.55$，隐藏层第 2 个神经元的输入 $\sum_i \omega_{i2}a_i=0.3925$，对 tanh 函数的导数为 $g'(\sum_i \omega_{i1}a_i)=1-g(\sum_i \omega_{i1}a_i)^2=0.860488$，$a_2=0.1$，因此：

$$\frac{\partial Loss_{o2}}{\omega_4}=-2(y_2-a_2)g'(\sum_j \omega_{j2}a_j)\omega_{j=2,k=2}g'(\sum_i \omega_{i2}a_i)a_2=\text{- }0.009525402681050708$$