DSAA 5002 - Data Mining and Knowledge Discovery in Data Science

(Fall Semester 2023)

Homework 2 Solutions

1.

(a) (10 marks) Try threshold 2, 5, and 8 for attributes A (that is, use the "A > 2, A < 2", "A > 5, A < 5", and "A > 8, A < 8" respectively). Use the Gini score to determine the best one θa among them.

Value 2: 0.4018
Value 5: 0.4844
Value 8: 0.4872
The best choice is 2.

(b) (15 marks) Use θa obtained above, and the Gini score, determine which attributes should firstly be used for developing a decision tree.

I(Parent) = 0.4922
Gini(A) = 0.4018. Gain(A) = 0.0904
Gini(B) = 0.4844, Gain(B) = 0.0078
Gini(C) = 0.4643, Gain(C) = 0.0279

① for "A>2, A<2"

	<2	>2
0	2	5
1	0	9

$I(A<2) = 1-(\frac{2}{2})^2 - 0 = 0$

$I(A>2) = 1-(\frac{5}{14})^2-(\frac{9}{14})^2 = \frac{45}{98}$

weighted Gini $= \frac{2}{16}\times 0 + \frac{14}{16}\times\frac{45}{98} = \frac{45}{112} \approx \boxed{0.4018}$

$\cancel{\Delta = \frac{63}{128} - \frac{45}{112} = \frac{81}{896} \approx 0.0904}$

② for "A>5, A<5"

	<5	>5
0	3	4
1	5	4

$I(A<5) = 1-(\frac{3}{8})^2 - (\frac{5}{8})^2 = \frac{15}{32}$

$I(A>5) = 1-(\frac{1}{2})^2-(\frac{1}{2})^2 = \frac{1}{2}$

weighted Gini $= \frac{1}{2}\times\frac{15}{32} + \frac{1}{2}\times\frac{1}{2} = \frac{31}{64} \approx \boxed{0.4844}$

$\cancel{\Delta = \frac{63}{128} - \frac{31}{64} = \frac{1}{128} \approx 0.0078}$

③ for "A>8, A<8"

	<8	>8
0	6	1
1	7	2

$I(A<8) = 1-(\frac{6}{13})^2-(\frac{7}{13})^2 = \frac{84}{169}$

$I(A>8) = 1-(\frac{1}{3})^2-(\frac{2}{3})^2 = \frac{4}{9}$

weighted Gini $= \frac{13}{16}\times\frac{84}{169} + \frac{3}{16}\times\frac{4}{9} = \frac{19}{39} \approx \boxed{0.4872}$

$\cancel{\Delta = \frac{63}{128} - \frac{19}{39} = \frac{25}{4992} \approx 0.0050}$

$\because Gini_2 < Gini_5 < Gini_8$

$\therefore$ The best threshold is 2 : "A>2, A<2".

① for attribute "A". from part (a) we know that

weighted Gini $= \frac{2}{16}Gini(2,0) + \frac{14}{16}Gini(5,9) = \frac{45}{112}$

② for attribute "B".

	Yes	No
0	4	3
1	4	5

$I(B==Yes) = 1-(\frac{4}{8})^2-(\frac{4}{8})^2 = \frac{1}{2}$

$I(B==No) = 1-(\frac{3}{8})^2-(\frac{5}{8})^2 = \frac{15}{32}$

Weighted Gini $= \frac{1}{2}\times\frac{1}{2} + \frac{1}{2}\times\frac{15}{32} = \frac{31}{64}$

③ for attribute "C"

	Yes	No
0	3	4
1	6	3

$I(C==Yes) = 1-(\frac{3}{9})^2-(\frac{6}{9})^2 = \frac{4}{9}$

$I(C==No) = 1-(\frac{4}{7})^2-(\frac{3}{7})^2 = \frac{24}{49}$

Weighted Gini $= \frac{9}{16}\times\frac{4}{9} + \frac{7}{16}\times\frac{24}{49} = \frac{13}{28}$

$\because I(parent) = 1-(\frac{7}{16})^2-(\frac{9}{16})^2 = \frac{63}{128}$

$\therefore \Delta_A = \frac{63}{128} - \frac{45}{112} = \frac{81}{896} \approx 0.0904$

$\Delta_B = \frac{63}{128} - \frac{31}{64} = \frac{1}{128} \approx 0.0078$

$\Delta_C = \frac{63}{128} - \frac{13}{28} = \frac{25}{896} \approx 0.0279$

$\because \Delta_A > \Delta_C > \Delta_B$

$\therefore$ A should firstly be used for developing a decision tree

2.

(a) (10 marks) Consider the procedures of building a decision tree with Gini score. If we plan only to use the attributes a3 and a5 to predict the decision d2, which attribute should we use first?

	yes	no
d_2 = yes	9	1
d_2 = no	4	6

Gini= $\frac{13}{20}\left[1-\left(\frac{9}{13}\right)^2-\left(\frac{4}{13}\right)^2\right]+\frac{7}{20}\left[1-\left(\frac{1}{7}\right)^2-\left(\frac{6}{7}\right)^2\right]=0.3626.$

If we use a5,

	yes	no
d_2 = yes	1	9
d_2 = no	3	7

Gini = $\frac{4}{20}\left[1-\left(\frac{1}{4}\right)^2-\left(\frac{3}{4}\right)^2\right]+\frac{16}{20}\left[1-\left(\frac{9}{16}\right)^2-\left(\frac{7}{16}\right)^2\right]=0.4688$

Therefore, a3 should be used in the first step.

(b) (20 marks) Use the naïve Bayes algorithm, the attributes a1 (with the threshold θ_1 = 37.95), a2, and a3 only, to predict the decision d2 for the following data of a new patient. (For simplicity you do NOT need to use the Laplacian correction.)

P(d2=yes) = 1/2
P(d2=no) = 1/2
P(a1>37.95|d2=yes) = 9/10
P(a1>37.95|d2=no) = 1/10
P(a2=yes|d2=yes) = 0
P(a2=yes|d2=no) = 1/10
P(a3=no|d2=yes) = 1/10
P(a3=no|d2=no) = 6/10

Using the naïve Bayes assumption,
P(d2=yes|a1>37.95, a2=no, a3=no) = 0
P(d2=no|a1>37.95, a2=no, a3=no) = 0.003

3.
(a) (7 marks) What is P(¬A, B, ¬C, D)?
P(¬A,B,¬C,D) = P(¬C|¬A) P(D|¬A,B) P(¬A) P(B) = (1 – 0.2)(0.6)(1 – 0.1)(0.5) = 0.216
(b) (8 marks) What is P(A | B, C, D) ?
P(A|B,C,D) = P(A,B,C,D) / P(B,C,D) P(A,B,C,D) = (0.7)(0.9)(0.1)(0.5) = 0.0315
P(B,C,D) = P(A,B,C,D) + P(¬A,B,C,D) = 0.0315 + (0.2)(0.6)(0.9)(0.5) = 0.0855
So, P(A | B,C,D) = 0.0315 / 0.0855 = 0.368

4.
(a) (5 marks)

$$y = \text{Relu}\left(\underset{1\times2}{V} \cdot \text{Relu}(\underset{2\times3}{W} \cdot \underset{3\times1}{x})\right)_{1\times1} \quad \text{scalar output}$$

set: $z = W \cdot x$

$a = \text{Relu}(z)$

$\therefore \quad y = \text{Relu}(V \cdot a)$

(c) (10 marks)

forward Pass $\quad X = (1,2,1)^T$

$$z = W \cdot X = \begin{pmatrix} 1 & 0 & 1 \\ -3 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -5 \end{pmatrix}$$

$$a = \text{ReLU}(W \cdot x) = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

$$y = \text{ReLU}(V \cdot a) = \text{ReLU}\left[(0 \ \ 1) \begin{pmatrix} 2 \\ 0 \end{pmatrix} \right] = 0$$

(d) (15 marks) $\hat{y}$ refers to o here

backward pass

$$\underset{1\times1}{J} = \frac{1}{2}(y - \hat{y})^2$$

$$y = \text{ReLU}(\underset{1\times2}{V} \cdot \underset{2\times1}{a})$$

$$\underset{2\times1}{a} = \text{ReLU}(\underset{2\times1}{z}) \ , \quad \underset{2\times1}{z} = \underset{2\times3}{W} \cdot \underset{3\times1}{X}$$

$$\frac{\partial J}{\partial V} = \frac{\partial J}{\partial y} \cdot \frac{\partial y}{\partial V} = \frac{\partial}{\partial y}\left[\frac{1}{2}(y-\hat{y})^2\right] \cdot \frac{\partial y}{\partial V}$$

$$= (y - \hat{y}) \ \text{ReLU}'(Va) \cdot a^T$$

$$= (1 - 0) \ \text{ReLU}'(0) \cdot (2 \ \ 0)$$

$$= (0 \ \ 0)$$