CS 5350/6350: Machine Learining Spring 2019

Homework 1
Xiang Zhang
ID: u1199149

Handed out: 25 January, 2019
Due: 11:59pm, 10 Feb, 2019

Decision Tree

1. First name the 7 instances as 1,2,$\cdots$, 7. Denote $\mathcal{S}, \mathcal{S}_0$ and $\mathcal{S}_1$ as the instance set at some node, the set with attribute value 0 and the set with attribute value 1, respectively. Then the info. gain is

$$Gain(\mathcal{S}, x) = H(\mathcal{S}) - \frac{|\mathcal{S}_0|}{|\mathcal{S}|} H(\mathcal{S}_0) - \frac{|\mathcal{S}_1|}{|\mathcal{S}|} H(\mathcal{S}_1)$$

   (a) For the first splitting, we have the following table:

attri.	$\mathcal{S}_0$	$\mathcal{S}_1$	(p_0^+, p_0^-)	(p_1^+, p_1^-)	$H(\mathcal{S}_0)$	$H(\mathcal{S}_1)$	$H(\mathcal{S})$	info. gain
x_1	$\{1,2,3,5,7\}$	$\{4,6\}$	$(0.2, 0.8)$	$(0.5, 0.5)$	0.72	1	0.86	0.06
x_2	$\{1,3,4\}$	$\{2,5,6,7\}$	$(\frac{2}{3}, \frac{1}{3})$	$(0, 1)$	0.92	0	0.86	0.47
x_3	$\{2,4,6,7\}$	$\{1,3,5\}$	$(0.25, 0.75)$	$(\frac{1}{3}, \frac{2}{3})$	0.81	0.92	0.86	0.03
x_4	$\{1,2,5,6\}$	$\{3,4,7\}$	$(0, 1)$	$(\frac{2}{3}, \frac{1}{3})$	0	0.92	0.86	0.47

   Since x_2 and x_4 have the same (largest) gain , we choose x_2 as the first attribute. In the second splitting, for $\mathcal{S} = \{1,3,4\}$ with $x_2 = 0$, we have

attri.	$\mathcal{S}_0$	$\mathcal{S}_1$	(p_0^+, p_0^-)	(p_1^+, p_1^-)	$H(\mathcal{S}_0)$	$H(\mathcal{S}_1)$	$H(\mathcal{S})$	info. gain
x_1	$\{1,3\}$	$\{4\}$	$(0.5, 0.5)$	$(1, 0)$	1	0	0.92	0.25
x_3	$\{4\}$	$\{1,3\}$	$(1, 0)$	$(0.5, 0.5)$	0	1	0.92	0.25
x_4	$\{1\}$	$\{3,4\}$	$(0, 1)$	$(1, 0)$	0	0	0.92	0.92

   from which we see that x_4 has the largest info. gain and hence x_4 is chosen as splitting attribute.

   In the second splitting, for $\mathcal{S} = \{2,5,6,7\}$ with $x_2 = 1$, since all the instances have the same label. The structure is shown in Fig. 1.

   (b) The function mapping is as follows ($(x_2, x_4) = (0, 1) \Rightarrow \text{label} = 1$).

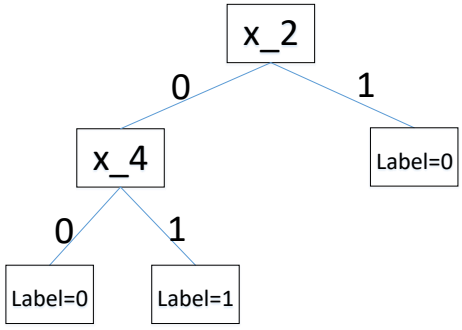

Figure 1: Decision tree for Problem 1 (a) using info. gain metric

x_1	x_2	x_3	x_4	y
0	0	0	0	0
0	0	0	1	1
0	0	1	0	0
0	0	1	1	1
0	1	0	0	0
0	1	0	1	0
0	1	1	0	0
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	0
1	0	1	1	1
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	1	0

2. (a) The majority error is defined as $ME = \min\{p^+, p^-\}$ (+ means 'play') and the gain is

$$Gain\_ME(\mathcal{S}, A) = ME(\mathcal{S}) - \sum_{v \in \text{Values}(A)} \frac{|\mathcal{S}_v|}{|\mathcal{S}|} ME(\mathcal{S}_v) \tag{1}$$

$$Gain\_GI(\mathcal{S}, A) = GI(\mathcal{S}) - \sum_{v \in \text{Values}(A)} \frac{|\mathcal{S}_v|}{|\mathcal{S}|} GI(\mathcal{S}_v) \tag{2}$$

▶ For 'Outlook', the spit subset is

$$
\begin{aligned}
S_{\text{sunny}} &= \{1,2,8,9,11\}, p^+ = 0.4, ME = 0.4, GI = 0.48 && (3)\\
S_{\text{overcast}} &= \{3,7,12,13\}, p^+ = 1, ME = 0, GI = 0 && (4)\\
S_{\text{rainy}} &= \{4,5,6,10,14\}, p^+ = 0.6, ME = 0.4, GI = 0.48 && (5)\\
Gain\_ME &= \frac{5}{14} - \frac{5}{14} \times (0.4 + 0.4) - \frac{4}{14} \times 0 = 0.07 && (6)\\
Gain\_GI &= 0.46 - \frac{5}{14} \times (0.48 + 0.48) = 0.12 && (7)
\end{aligned}
$$

▶ For 'Temperature', the spit subset is

$$
\begin{aligned}
S_{\text{hot}} &= \{1,2,3,13\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (8)\\
S_{\text{medium}} &= \{4,8,10,11,12,14\}, p^+ = \frac{2}{3}, ME = \frac{1}{3}, GI = 0.44 && (9)\\
S_{\text{cool}} &= \{5,6,7,9\}, p^+ = 0.75, ME = 0.25, GI = 0.38 && (10)\\
Gain\_ME &= \frac{5}{14} - \frac{4}{14} \times (0.5 + 0.25) - \frac{6}{14} \times \frac{1}{3} = 0 && (11)\\
Gain\_GI &= 0.46 - \frac{4 \times 0.5 + 6 \times 0.44 + 4 \times 0.38}{14} = 0.02 && (12)
\end{aligned}
$$

▶ For 'Humidity', the spit subset is

$$
\begin{aligned}
S_{\text{high}} &= \{1,2,3,4,8,12,14\}, p^+ = \frac{3}{7}, ME = \frac{3}{7}, GI = 0.49 && (13)\\
S_{\text{normal}} &= \{5,6,7,9,10,11,13\}, p^+ = \frac{6}{7}, ME = \frac{1}{7}, GI = 0.24 && (14)\\
S_{\text{low}} &= \emptyset && (15)\\
Gain\_ME &= \frac{5}{14} - \frac{7}{14} \times (\frac{3}{7} + \frac{1}{7}) = 0.07 && (16)\\
Gain\_GI &= 0.46 - \frac{7 \times 0.49 + 7 \times 0.24}{14} = 0.095 && (17)
\end{aligned}
$$

▶ For 'Wind', the spit subset is

$$
\begin{aligned}
S_{\text{strong}} &= \{2,6,7,11,12,14\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (18)\\
S_{\text{weak}} &= \{1,3,4,5,8,9,10,13\}, p^+ = 0.75, ME = 0.25, GI = 0.38 && (19)\\
Gain\_ME &= \frac{5}{14} - \frac{6}{14} \times 0.5 - \frac{8}{14} \times 0.25 = 0 && (20)\\
Gain\_GI &= 0.46 - \frac{6 \times 0.5 + 8 \times 0.38}{14} = 0.03 && (21)
\end{aligned}
$$

Since 'Outlook' has the largest gain (both ME & GI), we choose 'outlook' as the first attribute to split. The resulting subset is

$$
\begin{aligned}
S_{\text{sunny}} &= \{1,2,8,9,11\}, p^+ = 0.4, ME = 0.4, GI = 0.48 && (22)\\
S_{\text{overcast}} &= \{3,7,12,13\}, p^+ = 1, ME = 0, GI = 0 && (23)\\
S_{\text{rainy}} &= \{4,5,6,10,14\}, p^+ = 0.6, ME = 0.4, GI = 0.48 && (24)
\end{aligned}
$$

(1) For $S_{\text{sunny}} = \{1,2,8,9,11\}, ME = 0.4, GI = 0.48$, the gain is as follows.

▶ For 'Temperature', the spit subset is

$$
\begin{aligned}
S_{\text{hot}} &= \{1,2\}, p^+ = 0, ME = 0, GI = 0 && (25)\\
S_{\text{medium}} &= \{8,11\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (26)\\
S_{\text{cool}} &= \{9\}, p^+ = 1, ME = 0, GI = 0 && (27)\\
Gain\_ME &= 0.4 - \frac{2}{5} \times 0.5 = 0.2 && (28)\\
Gain\_GI &= 0.48 - \frac{2}{5} \times 0.5 = 0.28 && (29)
\end{aligned}
$$

▶ For 'Humidity', the spit subset is

$$
\begin{aligned}
S_{\text{high}} &= \{1,2,8\}, p^+ = 0, ME = 0, GI = 0 && (30)\\
S_{\text{normal}} &= \{9,11\}, p^+ = 1, ME = 0, GI = 0 && (31)\\
S_{\text{low}} &= \emptyset && (32)\\
Gain\_ME &= 0.4 && (33)\\
Gain\_GI &= 0.48 && (34)
\end{aligned}
$$

▶ For 'Wind', the spit subset is

$$
\begin{aligned}
S_{\text{strong}} &= \{2,11\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (35)\\
S_{\text{weak}} &= \{1,8,9\}, p^+ = \frac{1}{3}, ME = \frac{1}{3}, GI = 0.44 && (36)\\
Gain\_ME &= 0.4 - \frac{2}{5} \times 0.5 - \frac{3}{5} \times \frac{1}{3} = 0 && (37)\\
Gain\_GI &= 0.48 - \frac{2}{5} \times 0.5 - \frac{3}{5} \times 0.44 = 0.02 && (38)
\end{aligned}
$$

Since 'humidity' has the largest gain (both ME & GI), we choose it as the second attribute to split. The resulting subset are

$$
\begin{aligned}
S_{\text{sunny,high}} &= \{1,2,8\}, p^+ = 0 && (39)\\
S_{\text{sunny,normal}} &= \{9,11\}, p^+ = 1 && (40)
\end{aligned}
$$

from which we see that both the subsets have identical labels for the instances therein.
(2) $S_{\text{overcast}} = \{3,7,12,13\}, p^+ = 1, ME = 0$ already has identical labels for all the instances therein.
(3) For $S_{\text{rainy}} = \{4,5,6,10,14\}, p^+ = 0.6, ME = 0.4, GI = 0.48$, the gain is as follows.
▶ For 'Temperature', the spit subset is

$$
\begin{aligned}
S_{\text{hot}} &= \emptyset && (41)\\
S_{\text{medium}} &= \{4,10,14\}, p^+ = 0.67, ME = 0.33, GI = 0.44 && (42)\\
S_{\text{cool}} &= \{5,6\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (43)\\
Gain\_ME &= 0.4 - \frac{3 \times 0.33 + 2 \times 0.5}{5} = 0 && (44)\\
Gain\_Gini &= 0.48 - 0.6 * 0.44 - 0.4 * 0.5 = 0.16 && (45)
\end{aligned}
$$

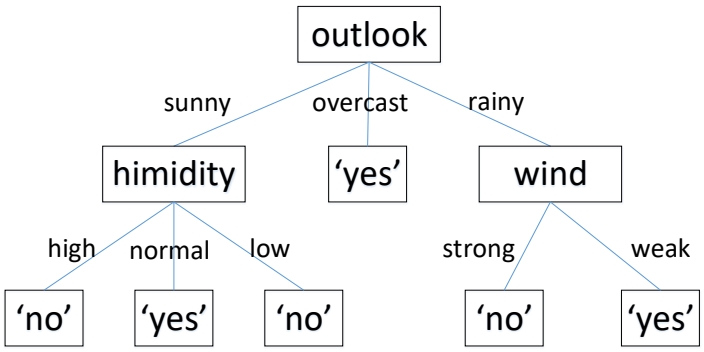

Figure 2: Decision tree for Problem 2 (a) using majority error metric

▶ For 'Humidity', the spit subset is

$$
\begin{aligned}
S_{\text{high}} &= \{4, 14\}, p^+ = 0.5, ME = 0.5, GI = 0.5 && (46)\\
S_{\text{normal}} &= \{5, 6, 10\}, p^+ = 0.67, ME = 0.33, Gi = 0.44 && (47)\\
S_{\text{low}} &= \emptyset && (48)\\
Gain\_ME &= 0.4 - \frac{2 \times 0.5 + 3 \times 0.33}{5} = 0 && (49)\\
Gain\_GI &= 0.48 - 0.4 * 0.5 - 0.6 * 0.44 = 0.16 && (50)
\end{aligned}
$$

▶ For 'Wind', the spit subset is

$$
\begin{aligned}
S_{\text{strong}} &= \{6, 14\}, p^+ = 0, ME = 0, GI = 0 && (51)\\
S_{\text{weak}} &= \{4, 5, 10\}, p^+ = 1, ME = 0, GI = 0 && (52)\\
Gain\_ME &= 0.4 && (53)\\
Gain\_GI &= 0.48 && (54)
\end{aligned}
$$

Since 'wind' has the largest gain (both ME & GI) we choose it as split attribute, and the resulting subset are

$$
\begin{aligned}
S_{\text{rainy,strong}} &= \{6, 14\}, p^+ = 0 && (55)\\
S_{\text{rainy,weak}} &= \{4, 5, 10\}, p^+ = 1 && (56)
\end{aligned}
$$

and all the instances within the same subset have identical labels.

The tree structure is shown in Fig. (2)

(b) The Gini index is computed in part (a). Since the two metrics agree, the tree structure is exactly the same as that of part (a).

(c) The trees built here and the tree in the lecture are exactly the same. The only possible difference is that if we choose 'humidity' (having the same ME gain as 'outlook') as the first splitting attribute, the tree structure might be different.

3. (a) The 15-th instance is (S, M, N, W, Yes). We calculate the info. gain as follows (Denote H as the expected entropy):

▶ For 'Outlook', we have

$$\begin{aligned} S_{\text{sunny}} &= \{1,2,8,9,11,15\}, p^+ = 0.5, H = 1 && (57)\\ S_{\text{overcast}} &= \{3,7,12,13\}, p^+ = 1, H = 0 && (58)\\ S_{\text{rainy}} &= \{4,5,6,10,14\}, p^+ = 0.6, H = 0.97 && (59)\\ IG &= 0.92 - \frac{6*1+4*0+5*0.97}{15} = 0.197 && (60) \end{aligned}$$

▶ For 'Temperature', the spit subset is

$$\begin{aligned} S_{\text{hot}} &= \{1,2,3,13\}, p^+ = 0.5, H = 1 && (61)\\ S_{\text{medium}} &= \{4,8,10,11,12,14,15\}, p^+ = \frac{5}{7}, H = 0.86 && (62)\\ S_{\text{cool}} &= \{5,6,7,9\}, p^+ = 0.75, H = 0.81 && (63)\\ IG &= 0.92 - \frac{4+7*0.86+4*0.81}{15} = 0.036 && (64) \end{aligned}$$

▶ For 'Humidity', the spit subset is

$$\begin{aligned} S_{\text{high}} &= \{1,2,3,4,8,12,14\}, p^+ = \frac{3}{7}, H = 0.985 && (65)\\ S_{\text{normal}} &= \{5,6,7,9,10,11,13,15\}, p^+ = \frac{7}{8}, H = 0.54 && (66)\\ S_{\text{low}} &= \emptyset && (67)\\ IG &= 0.92 - \frac{7*0.985+8*0.54}{15} = 0.17 && (68) \end{aligned}$$

▶ For 'Wind', the spit subset is

$$\begin{aligned} S_{\text{strong}} &= \{2,6,7,11,12,14\}, p^+ = 0.5, H = 1 && (69)\\ S_{\text{weak}} &= \{1,3,4,5,8,9,10,13,15\}, p^+ = \frac{7}{9}, H = 0.76 && (70)\\ IG &= 0.92 - \frac{6*1+9*0.76}{15} = 0.06 && (71) \end{aligned}$$

It can be seen that 'Outlook' has the largest IG, and thus is the best feature to split on.

(b) The 15-th instance in this case is (O, M, N, W, Yes). ▶ For 'Outlook', we have

$$\begin{aligned} S_{\text{sunny}} &= \{1,2,8,9,11\}, p^+ = 0.4, H = 0.97 && (72)\\ S_{\text{overcast}} &= \{3,7,12,13,15\}, p^+ = 1, H = 0 && (73)\\ S_{\text{rainy}} &= \{4,5,6,10,14\}, p^+ = 0.6, H = 0.97 && (74)\\ IG &= 0.92 - \frac{5*0.97+5*0.97}{15} = 0.27 && (75) \end{aligned}$$

and the IGs for the remaining features stay the same as part (a). The best feature is still 'Outlook'.

(c) Using the fractional count, the 15-th instance in this case is $(Outlook, M, N, W, \text{Yes})$ where $Outlook = \{\frac{5}{14}Sunny, \frac{4}{14}Overcast, \frac{5}{14}Rainy\}$

▶ For 'Outlook', we have

$$S_{\text{sunny}} = \{1,2,8,9,11,15(\frac{5}{14})\}, p^+ = \frac{\frac{5}{14}+2}{5+\frac{5}{14}} = 0.44, H = 0.99 \quad (76)$$

$$S_{\text{overcast}} = \{3,7,12,13,15(\frac{4}{15})\}, p^+ = 1, H = 0 \quad (77)$$

$$S_{\text{rainy}} = \{4,5,6,10,14,15(\frac{5}{15})\}, p^+ = \frac{3+\frac{5}{14}}{5+\frac{5}{14}} = 0.63, H = 0.95 \quad (78)$$

$$IG = 0.92 - \frac{5.36 * 0.99 + 5.36 * 0.95}{15} = 0.23 \quad (79)$$

implying the best feature is still 'Outlook'.

(d) Form part (c), we have 'outlook' being the first split feature and the resulting data subsets are

$$S_{\text{sunny}} = \{1,2,8,9,11,15(\frac{5}{14})\}, H = 0.99 \quad (80)$$

$$S_{\text{overcast}} = \{3,7,12,13,15(\frac{4}{15})\}, p^+ = 1, H = 0 \quad (81)$$

$$S_{\text{rainy}} = \{4,5,6,10,14,15(\frac{5}{15})\}, H = 0.95 \quad (82)$$

(1) For $S_{\text{sunny}} = \{1,2,8,9,11,15(\frac{5}{14})\}, H = 0.99$, the IGs are calculated as follows.

▶ 'Temperature'

$$S_{\text{hot}} = \{1,2\}, p^+ = 0, H = 0 \quad (83)$$

$$S_{\text{medium}} = \{8,11,15(\frac{5}{14})\}, p^+ = \frac{1+\frac{5}{14}}{2+\frac{5}{14}} = 0.58, H = 0.98 \quad (84)$$

$$S_{\text{cool}} = \{9\}, p^+ = 1, H = 0 \quad (85)$$

$$IG = 0.99 - \frac{2.36 * 0.98}{5.36} = 0.56 \quad (86)$$

▶ 'Humidity'

$$S_{\text{high}} = \{1,2,8\}, p^+ = 0, H = 0 \quad (87)$$

$$S_{\text{normal}} = \{9,11,15(\frac{5}{14})\}, p^+ = 1, H = 0 \quad (88)$$

$$S_{\text{low}} = \emptyset \quad (89)$$

$$IG = 0.99 \quad (90)$$

▶ 'Wind'

$$S_{\text{strong}} = \{2,11\}, p^+ = 0.5, H = 1 \quad (91)$$

$$S_{\text{normal}} = \{1,8,9,15(\frac{5}{14})\}, p^+ = \frac{1+\frac{5}{14}}{3+\frac{5}{14}} = 0.4, H = 0.97 \quad (92)$$

$$S_{\text{low}} = \emptyset \quad (93)$$

$$IG = 0.99 - \frac{2 * 1 + 3.36 * 0.97}{5.36} = 0.009 \quad (94)$$

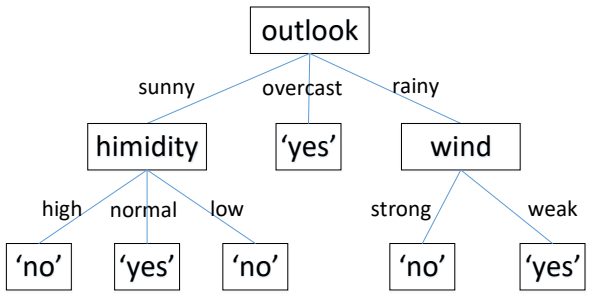

Figure 3: Decision tree for Problem 2 (a) using majority error metric

Clearly, 'humidity' is the best feature for splitting and the resulting subsets have unique labels.

(2) All the instances in $S_{\text{overcast}} = \{3, 7, 12, 13, 15(\frac{4}{15})\}$ have the same label, implying there is no need for further splitting.

(3) For $S_{\text{rainy}} = \{4, 5, 6, 10, 14, 15(\frac{5}{15})\}, H = 0.95$, the IGs are as follows.

▶ 'Temperature'

$$S_{\text{hot}} = \emptyset \tag{95}$$

$$S_{\text{medium}} = \{4, 10, 14, 15(\frac{5}{14})\}, p^+ = \frac{2 + \frac{5}{14}}{3 + \frac{5}{14}} = 0.7, H = 0.88 \tag{96}$$

$$S_{\text{cool}} = \{5, 6\}, p^+ = 0.5, H = 1 \tag{97}$$

$$IG = 0.95 - \frac{3.36 * 0.88 + 2 * 1}{5.36} = 0.025 \tag{98}$$

▶ 'Humidity'

$$S_{\text{high}} = \{4, 14\}, p^+ = 0.5, H = 1 \tag{99}$$

$$S_{\text{normal}} = \{5, 6, 10, 15(\frac{5}{14})\}, p^+ = \frac{2 + \frac{5}{14}}{3 + \frac{5}{14}} = 0.44, H = 0.88 \tag{100}$$

$$S_{\text{low}} = \emptyset \tag{101}$$

$$IG = 0.95 - \frac{2 + 3.36 * 0.99}{5.36} = 0.025 \tag{102}$$

▶ 'Wind'

$$S_{\text{strong}} = \{6, 14\}, p^+ = 0, H = 0 \tag{103}$$

$$S_{\text{normal}} = \{4, 5, 10, 15(\frac{5}{14})\}, p^+ = 1, H = 0 \tag{104}$$

$$S_{\text{low}} = \emptyset \tag{105}$$

$$IG = 0.95 \tag{106}$$

Clearly, 'Wind' is the best feature. The tree structure is shown in Fig. (3), which is the same as the tree in Fig. (2).

4. Show that the info, gain is always non-negative.

   Proof: The info. gain is defined as

$$IG = H(\mathbf{p}) - \sum_{i=1}^{V} r_i H(\mathbf{q}^i)$$

   in which $\mathbf{p} := (p_1, p_2, \cdots, p_L)$ is the label distribution of S and $\mathbf{q}^i := (q_1^i, q_2^i, \cdots, q_L^i)$ is the label distribution of S_i, and r_i denotes the sample portion $\frac{|S_i|}{S}$ satisfying $\sum_{i=1}^{V} r_i = 1$. L is the number of labels. The condition $\sum_{i=1}^{V} r_i q_\ell^i = p_\ell, \forall \ell \in [L]$ holds.

   Note that the function $f(x) := x \log x$ is convex when $x > 0$ since $f''(x) = \frac{1}{x} > 0$.

$$\begin{aligned}
IG &= H(\mathbf{p}) - \sum_{i=1}^{V} r_i H(\mathbf{q}^i) && (107)\\
&= \sum_{\ell=1}^{L} -p_\ell \log p_\ell - \sum_{i=1}^{V} \sum_{\ell=1}^{L} -r_i q_\ell^i \log q_\ell^i && (108)\\
&= \sum_{\ell=1}^{L} \left(\sum_{i=1}^{V} r_i q_\ell^i \log q_\ell^i - p_\ell \log p_\ell \right) && (109)\\
&= \sum_{\ell=1}^{L} \left(\sum_{i=1}^{V} r_i f(q_\ell^i) - f\left(\sum_{i=1}^{V} r_i q_\ell^i \right) \right) && (110)\\
&\overset{(1)}{\geq} 0 && (111)
\end{aligned}$$

   where (1) is due to the convexity of f, which implies $\sum_{i=1}^{V} r_i f(q_\ell^i) - f\left(\sum_{i=1}^{V} r_i q_\ell^i\right) \geq 0, \forall \ell \in [L]$. IG is actually the mutual information in information theory.

5. The **variance** might be used to measure the impurity of a data set. For $S = \{s_1, s_2, ..., s_N\}, s_i \in \mathbb{R}$, the variance is $\mathrm{Var}(S) = \sum_{i=1}^{N} (s_i - \bar{s})^2$, $\bar{s} = \frac{1}{N} \sum_{i=1}^{N} s_i$ and the gain defined as

$$Gain(S, A) = \mathrm{Var}(S) - \sum_{v \in \mathrm{Values(A)}} \frac{|S_v|}{|S|} \mathrm{Var}(S_v)$$

   We may also use the **absolute error** defined as $\mathrm{Abs}(S) = \sum_{i=1}^{N} |s_i - \bar{s}|$ and the gain is

$$Gain(S, A) = \mathrm{Abs}(S) - \sum_{v \in \mathrm{Values(A)}} \frac{|S_v|}{|S|} \mathrm{Abs}(S_v)$$

Decision Tree Practice

1. The Github repository link:

   `https://github.com/xiangzhang-122/Machine_Learning`

2. (a) The code of training is `tree_build_test.py`.

   (b) The prediction errors on training dataset are

depth|metric	gini_index	info_gain	ME
1	30.2%	30.2%	30.2%
2	22.2%	22.2%	30.1%
3	17.6%	18.1%	24.7%
4	8.9%	8.2%	21.3%
5	2.7%	2.7%	18.1%
6	0.0%	0.0%	17.2%

The prediction errors on testing dataset are

depth|metric	gini_index	info_gain	ME
1	29.7%	29.7%	29.7%
2	22.3%	22.3%	31.6%
3	18.4%	19.6%	26.2%
4	13.3%	14.7%	26.0%
5	9.5%	9.5%	23.1%
6	9.5%	9.5%	23.1%

(c) Training error becomes smaller and smaller as we increase the tree depth and vanishes if depth ≥ 6 for both info_gain and gini_index. However, after some point, the testing error will not decrease anymore even we increase the tree depth. This implies a tradeoff between training accuracy and testing accuracy (result of overfitting). We can also see that the majority error (ME) metric has the worst performance.

3. (a) By binary quantization of the numeric attribute values, we get the training errors (maximum depth is less than 16)

depth|metric	gini_index	info_gain	ME
1	10.88%	11.92%	10.88%
2	10.42%	10.60%	10.42%
3	9.34%	10.06%	9.60%
4	7.48%	7.92%	8.32%
5	5.96%	6.12%	7.06%
6	4.68%	4.72%	6.72%
7	3.46%	3.48%	6.44%
8	2.66%	2.86%	6.28%
9	2.12%	2.30%	6.22%
10	1.70%	1.70%	6.20%
11	1.46%	1.44%	6.18%
12	1.38%	1.36%	6.18%
13	1.36 %	1.36%	6.18%

The testing errors are

depth|metric	gini_index	info_gain	ME
1	11.66%	12.48%	11.66%
2	10.88%	11.14%	10.89%
3	11.50%	10.76%	11.34%
4	11.94%	11.88%	11.70%
5	12.78%	12.28%	11.48%
6	14.14%	13.20%	11.80%
7	15.18%	14.24%	11.92%
8	16.26%	14.90%	11.98%
9	16.54%	15.52%	11.98%
10	16.72%	15.92%	12.00%
11	17.08%	15.93%	12.00%
12	17.06%	15.92%	
13	17.04%	15.96%	

(b) By fulfilling the missing values by the majority value of the corresponding attributes, we get the training errors

depth|metric	gini_index	info_gain	ME
1	10.88%	11.92%	10.88%
2	10.52%	10.6%	10.50%
3	10.10%	10.22%	9.76%
4	8.76%	8.68%	8.64%
5	7.38%	7.14%	7.84%
6	5.72%	5.68%	7.48%
7	4.50%	4.52%	7.28%
8	3.70%	3.86%	7.20%
9	2.94%	3.20%	7.18%
10	2.46%	2.64%	7.18%
11	2.24%	2.34%	
12	2.20%	2.22%	
13	2.20%	2.20%	

and the testing errors

depth|metric	gini_index	info_gain	ME
1	11.66%	12.48%	11.66%
2	11.04%	11.14%	11.02%
3	10.84%	10.76%	11.54%
4	11.64%	11.74%	11.54%
5	12.26%	12.2%	11.54%
6	13.26%	13.48%	11.92%
7	14.30%	14.14%	11.94%
8	15.48%	15.18%	12.00%
9	16.00%	15.74%	12.04%
10	16.24%	16.04%	12.04%
11	16.56%	16.44%	
12	16.52%	16.34%	
13	16.54%	16.36%	

(c) We have the following observations: 1) For training dataset, the training error becomes smaller and smaller as we increase the depth of the tree. However, for testing dataset, as the tree depth goes up, the test error goes down a little bit first and then goes up, which reflects the effect of overfitting of the training data. 2) The training error is generally smaller than the test error. 3) For the training error, treating 'unknown' as a new attribute actually results in slightly smaller error than the majority values replacement.