

# Homework 3 - Classification; Decision Tree; Ensemble Learning

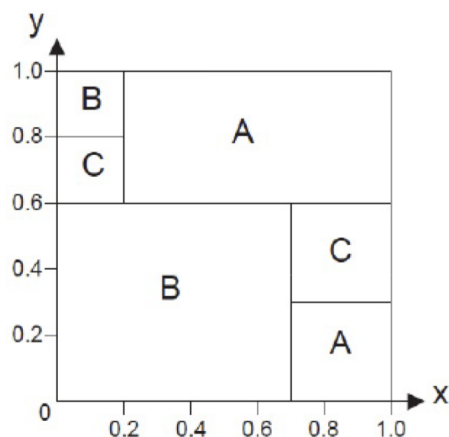
Benny Chen

April 11, 2023

## 1 Classification and Decision Tree

### 1.1 General Knowledge about Decision Tree

Consider a training set sampled uniformly from the two-dimensional space shown in the following figure.



The space is divided into three classes A, B, and C. Assume that the training set size is large enough so that the probabilities can be calculated accurately based on the areas of the selected regions (for instance, we can find the probability of sampling a data point of class B is  $p(B) = 0.42 + 0.04$ ). You do not need to build a tree here, but to calculate some essential entropy values to prepare a tree.

- (a) Please compute the entropy for the overall data (considering that we have not decide any split point).
- (b) If we split the data at  $x \leq 0.7$ , what is the resulting entropy value?

(a) We first calculate the areas of each class:

$$p(A) = (0.3 * 0.3) + (0.8 * 0.4) = 0.41 \quad (1)$$

$$p(B) = (0.2 * 0.2) + (0.6 * 0.7) = 0.46 \quad (2)$$

$$p(C) = (0.2 * 0.2) + (0.3 * 0.3) = 0.13 \quad (3)$$

We then calculate the entropy of each class:

$$-0.41 * \log_2(0.41) - 0.46 * \log_2(0.46) - 0.13 * \log_2(0.13) = 1.425 \quad (4)$$

(b) We first calculate the areas of each class:

$$p(A) = (0.5 * 0.4) = 0.2 \quad (5)$$

$$p(B) = (0.2 * 0.2) + (0.6 * 0.7) = 0.46 \quad (6)$$

$$p(C) = (0.2 * 0.2) = 0.04 \quad (7)$$

We then calculate the entropy of each class:

$$-0.2 * \log_2(0.2) - 0.46 * \log_2(0.46) - 0.04 * \log_2(0.04) = 1.16 \quad (8)$$

We then multiply the entropy by the split which is 0.7:

$$0.7 * 1.16 = 0.812 \quad (9)$$

We also can calculate the entropy of the other half:

$$p(A) = (0.3 * 0.3) + (0.3 * 0.4) = 0.21 \quad (10)$$

$$p(C) = (0.3 * 0.3) = 0.09 \quad (11)$$

We then calculate the entropy of each class:

$$-0.21 * \log_2(0.21) - 0.09 * \log_2(0.09) = 0.785 \quad (12)$$

We then multiply the entropy by the split which is 0.3:

$$0.3 * 0.785 = 0.235 \quad (13)$$

We then add the two entropy values together:

$$0.812 + 0.235 = 1.047 \quad (14)$$

ID	Maximum Health	Stealth Skill	Listen Mode Distance	Clear the Episode
1	Enough	Medium	A	Yes
2	High	Medium	A	Yes
3	High	Strong	B	Yes
4	High	Strong	C	Yes
5	Few	Medium	B	No
6	High	Medium	A	No
7	Few	Strong	C	No
8	Few	Strong	A	No

## 1.2

Suppose you are a player of THE LAST of US or imagine you are Joel or Ellie in this apocalypse. The following shows a history of a player in trying one episode of this game. We also indicate whether the episode is cleared or not in the last column. Note that the first column "ID" is for us to refer the record number only.

- (i) Please find a C4.5 decision tree according to the above example. In the decision tree, whenever we process (1) a node containing at least 80% records with the same label or (2) a node containing at most 2 records, we stop to process this node for splitting.
- (ii) Consider a trial with high maximum health, medium stealth skill, and A-level listen mode distance. Please estimate the probability that this trial will clear the episode.
- (i) Entropy(Clear the Episode):

$$-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 1 \quad (15)$$

Entropy(Maximum Health):

$$Entropy_{High} = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 0.811 \quad (16)$$

$$Entropy_{Few} = -1\log_2(1) = 0 \quad (17)$$

$$Entropy_{Enough} = -1\log_2(1) = 0 \quad (18)$$

Entropy(Clear the Episode — Maximum Health):

$$Entropy = \frac{4}{8}Entropy_{High} + \frac{3}{8}Entropy_{Few} + \frac{1}{8}Entropy_{Enough} = 0.405 \quad (19)$$

Split Information(Maximum Health):

$$SplitInfo = -\frac{4}{8}\log_2\left(\frac{4}{8}\right) - \frac{3}{8}\log_2\left(\frac{3}{8}\right) - \frac{1}{8}\log_2\left(\frac{1}{8}\right) = 1.40563 \quad (20)$$

Gain(Maximum Health):

$$Gain = (1 - 0.405)/1.40563 = 0.424 \quad (21)$$

Entropy(Stealth Skill):

$$Entropy_{Medium} = -\frac{2}{4}\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4}) = 1 \quad (22)$$

$$Entropy_{Strong} = -\frac{2}{4}\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4}) = 1 \quad (23)$$

Entropy(Clear the Episode — Stealth Skill):

$$Entropy = \frac{4}{8}Entropy_{Medium} + \frac{4}{8}Entropy_{Strong} = 1 \quad (24)$$

Split Information(Stealth Skill):

$$SplitInfo = -\frac{4}{8}\log_2(\frac{4}{8}) - \frac{4}{8}\log_2(\frac{4}{8}) = 1 \quad (25)$$

Gain(Stealth Skill):

$$Gain = (1 - 1)/1 = 0 \quad (26)$$

Entropy(Listen Mode Distance):

$$Entropy_A = -\frac{2}{4}\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4}) = 1 \quad (27)$$

$$Entropy_B = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1 \quad (28)$$

$$Entropy_C = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1 \quad (29)$$

Entropy(Clear the Episode — Listen Mode Distance):

$$Entropy = \frac{4}{8}Entropy_A + \frac{2}{8}Entropy_B + \frac{2}{8}Entropy_C = 1 \quad (30)$$

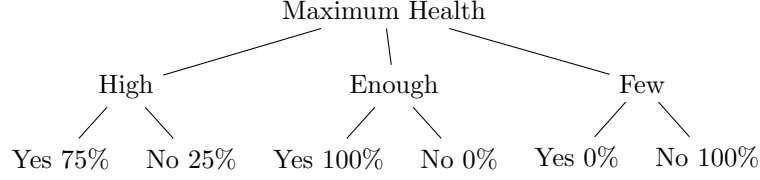
Split Information(Listen Mode Distance):

$$SplitInfo = -\frac{4}{8}\log_2(\frac{4}{8}) - \frac{2}{8}\log_2(\frac{2}{8}) - \frac{2}{8}\log_2(\frac{2}{8}) = 1.5 \quad (31)$$

Gain(Listen Mode Distance):

$$Gain = (1 - 1)/1.5 = 0 \quad (32)$$

From the above calculation, we can see that the maximum gain is from Maximum Health, so we choose Maximum Health as the first node. We can now start creating the decision tree.



We now have to continue splitting the tree. We will use the same method as above to calculate the gain but this time only for Maximum Health = High data points so ID's 2,3,4, and 6.

Entropy(Clear the Episode):

$$-\frac{3}{4}\log_2(\frac{3}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) = 0.811 \quad (33)$$

Entropy(Stealth Skill):

$$Entropy_{Medium} = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1 \quad (34)$$

$$Entropy_{Strong} = -1\log_2(1) - 0\log_2(0) = 0 \quad (35)$$

Entropy(Clear the Episode — Stealth Skill):

$$Entropy = \frac{2}{4}Entropy_{Medium} + \frac{2}{4}Entropy_{Strong} = 0.5 \quad (36)$$

Split Information(Stealth Skill):

$$SplitInfo = -\frac{2}{4}\log_2(\frac{2}{4}) - \frac{2}{4}\log_2(\frac{2}{4}) = 1 \quad (37)$$

Gain(Stealth Skill):

$$Gain = (0.811 - 0.5)/1 = 0.311 \quad (38)$$

Entropy(Listen Mode Distance):

$$Entropy_A = -\frac{1}{2}\log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2}) = 1 \quad (39)$$

$$Entropy_B = -1\log_2(1) - 0\log_2(0) = 0 \quad (40)$$

$$Entropy_C = -1\log_2(1) - 0\log_2(0) = 0 \quad (41)$$

Entropy(Clear the Episode — Listen Mode Distance):

$$Entropy = \frac{2}{4}Entropy_A + \frac{1}{4}Entropy_B + \frac{1}{4}Entropy_C = 0.5 \quad (42)$$

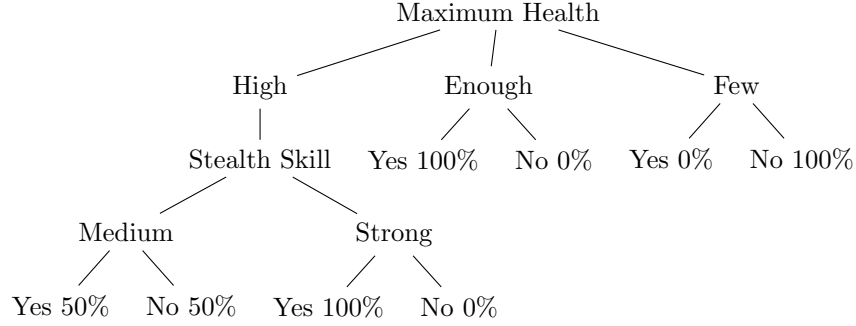
Split Information(Listen Mode Distance):

$$SplitInfo = -\frac{2}{4}\log_2(\frac{2}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) - \frac{1}{4}\log_2(\frac{1}{4}) = 1.5 \quad (43)$$

Gain(Listen Mode Distance):

$$Gain = (0.811 - 0.5)/1.5 = 0.207 \quad (44)$$

The maximum gain is from Stealth Skill, so we choose Stealth Skill as the second node. We can now continue creating the decision tree.



This is the final decision tree as we have no more data points to split on as after Medium there is the Listen Mode Distance however the given data shows that A is 50% Yes and 50% No. We can now use this decision tree to predict whether a player will clear the episode or not.

- (ii) Given a trail of a HIGH maximum health, a MEDIUM stealth skill and a A listen mode distance we can predict that the player has a 50% chance of clearing the episode.

### 1.3

The following shows a history of customers with their incomes, ages and an attribute called "Have iPhone" indicating whether they have an iPhone. We also indicate whether they will buy an iPad or not in the last column.

ID	Income	Age	Have.iPhone	Buy.iPad
1	High	Young	Yes	Yes
2	High	Old	Yes	Yes
3	Medium	Young	No	Yes
4	High	Old	No	Yes
5	Medium	Young	No	No
6	Medium	Young	No	No
7	Medium	Old	No	No
8	Medium	Old	No	No

We want to train a CART decision tree classifier (based on gini index) to predict whether a new customer will buy an iPad or not. We define the value of attribute "Buy iPad" is the label of a record.

- (i) Please find a CART decision tree according to the above example. In the decision tree, whenever we process a node containing at most 3 records, we stop to process this node for splitting.
- (ii) Consider a new young customer whose income is medium and he has an iPhone. Please predict whether this new customer will buy an iPad or not.
- (i) Gini Index(Income):

$$Gini_{High} = 1 - (3/3)^2 - (0/3)^2 = 0 \quad (45)$$

$$Gini_{Medium} = 1 - (1/5)^2 - (4/5)^2 = 0.32 \quad (46)$$

Sum Gini Index(Income):

$$Gini_{Income} = \frac{3}{8}Gini_{High} + \frac{5}{8}Gini_{Medium} = 0.2 \quad (47)$$

Gini Index(Age):

$$Gini_{Young} = 1 - (2/4)^2 - (2/4)^2 = 0.5 \quad (48)$$

$$Gini_{Old} = 1 - (2/4)^2 - (2/4)^2 = 0.5 \quad (49)$$

Sum Gini Index(Age):

$$Gini_{Age} = \frac{4}{8}Gini_{Young} + \frac{4}{8}Gini_{Old} = 0.5 \quad (50)$$

Gini Index(Have iPhone):

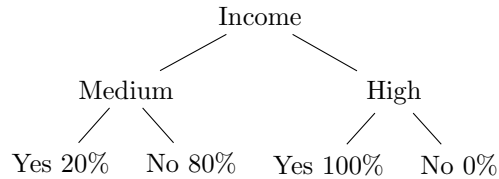
$$Gini_{Yes} = 1 - (2/2)^2 - (0/2)^2 = 0 \quad (51)$$

$$Gini_{No} = 1 - (2/6)^2 - (4/6)^2 = 0.44 \quad (52)$$

Sum Gini Index(Have iPhone):

$$Gini_{HaveiPhone} = \frac{2}{8}Gini_{Yes} + \frac{6}{8}Gini_{No} = 0.33 \quad (53)$$

We can see that the minimum Gini Index is from Income, so we choose Income as the first node. We can now continue creating the decision tree.



We can now keep splitting the tree but now we only need to look at the Medium branch as the High branch is already 100% Yes. We can now

look at the Medium branch with ID's of 3, 5, 6, 7 and 8. We can now calculate the Gini Index for Age and Have iPhone.

Gini Index(Age):

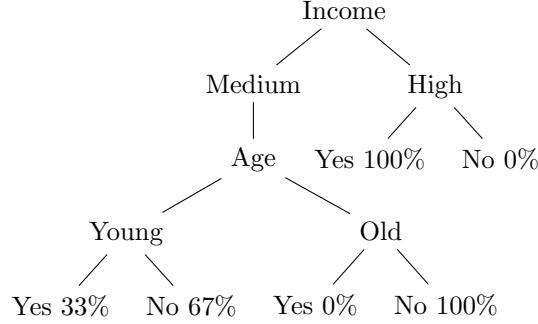
$$Gini_{Young} = 1 - (1/3)^2 - (2/3)^2 = 0.44 \quad (54)$$

$$Gini_{Old} = 1 - (2/2)^2 - (0/2)^2 = 0 \quad (55)$$

Sum Gini Index(Age):

$$Gini_{Age} = \frac{3}{5}Gini_{Young} + \frac{2}{5}Gini_{Old} = 0.26 \quad (56)$$

Medium Income also all has the same value for Have iPhone so we can't split on that. We can now create the decision tree.



We can now stop as we have reached the minimum number of records.

- (ii) Given the question to predict if a Medium Income, Young Age and Have iPhone will buy an iPad or not. We can see that the answer is 67% No and 33% Yes. So we can predict that the customer will not buy an iPad.

## 2 Ensemble Learning

Data Point $i$	1	2	3	4	5	6	7	8
Weight, $w_i^{(j)}$	0.35	0.2	0.1	0.05	0.05	0.05	0.1	0.1
$y_i$	-1	+1	+1	+1	+1	-1	-1	-1
$f(x_i)$	+1	+1	+1	+1	-1	-1	-1	-1

Consider the above 8 weighted training examples along with their ground-truths ( $y_i$ ) and predicted class labels ( $f(X_i)$ ) after performing  $j$  iterations or rounds of boosting iterations.

1. Based on the given information, calculate  $\epsilon_j$  and  $\alpha_j$ .



2. Show the new weights for each of the 8 training examples in the next iteration  $W^{(j+1)}$

1. Error Rate Equation:

$$\epsilon_j = \frac{1}{N} \sum_{i=1}^N w_i^{(j)} \delta(y_i \neq f(x_i)) \quad (57)$$

Error Rate for given data:

$$\epsilon_j = \frac{0.35 + 0.05}{8} = 0.05 \quad (58)$$

$\alpha_j$  Equation:

$$\alpha_j = \frac{1}{2} \ln \left( \frac{1 - \epsilon_j}{\epsilon_j} \right) \quad (59)$$

$\alpha_j$  for given data:

$$\alpha_j = \frac{1}{2} \ln \left( \frac{1 - 0.05}{0.05} \right) = 1.47 \quad (60)$$

2. New Weight Equation:

$$w_i^{(j+1)} = w_i^{(j)} \begin{cases} e^{-\alpha_j} & \text{if } y_i = f(x_i) \\ e^{\alpha_j} & \text{if } y_i \neq f(x_i) \end{cases} \quad (61)$$

New Weight for given data:

$$w_1^{(j+1)} = 0.35e^{1.47} = 1.52 \quad (62)$$

$$w_2^{(j+1)} = 0.2e^{-1.47} = 0.04 \quad (63)$$

$$w_3^{(j+1)} = 0.1e^{-1.47} = 0.02 \quad (64)$$

$$w_4^{(j+1)} = 0.05e^{-1.47} = 0.01 \quad (65)$$

$$w_5^{(j+1)} = 0.05e^{1.47} = 0.21 \quad (66)$$

$$w_6^{(j+1)} = 0.05e^{-1.47} = 0.01 \quad (67)$$

$$w_7^{(j+1)} = 0.1e^{-1.47} = 0.02 \quad (68)$$

$$w_8^{(j+1)} = 0.1e^{-1.47} = 0.02 \quad (69)$$