Temp	Midterm	Homework_done	Quiz	Go_BBQ
<15°C	High	No	Bad	No
<15°C	High	No	Good	No
16-25°C	High	No	Bad	Yes
>26°C	Medium	No	Bad	Yes
>26°C	Low	Yes	Bad	Yes
>26°C	Low	Yes	Good	No
16-25°C	Low	Yes	Good	Yes
<15°C	Medium	No	Bad	No
<15°C	Low	Yes	Bad	Yes
>26°C	Medium	Yes	Bad	Yes
<15°C	Medium	Yes	Good	Yes
16-25°C	Medium	No	Good	Yes
16-25°C	High	Yes	Bad	Yes
<15°C	Medium	No	Good	No

```python
import math
# import numpy as np

total_n = 14
class1n = 6
class2n = 4
class3n = 4

class1y = 2
class2y = 0
class3y = 4

def cal_en(total_n, class1n, class2n, class3n, class1y, class2y, class3y):
    total_y = class1y + class2y + class3y
    en = -(((total_y/total_n) * math.log2(total_y/total_n + 1e-8)) + (((total_n-total_y)/total_n) * math.log2((total_n-total_y)/total_n + 1e-8)))
    en1 = -(((class1y/class1n) * math.log2(class1y/class1n + 1e-8)) + (((class1n-class1y)/class1n) * math.log2((class1n-class1y)/class1n + 1e-8)))
    en2 = -(((class2y/class2n) * math.log2(class2y/class2n + 1e-8)) + (((class2n-class2y)/class2n) * math.log2((class2n-class2y)/class2n + 1e-8)))

    if class3n == 0:
        en3 = 0
        gain = en - class1n/total_n * en1 - class2n/total_n * en2
    else:
        en3 = -(((class3y/class3n) * math.log2(class3y/class3n + 1e-8)) + (((class3n-class3y)/class3n) * math.log2((class3n-class3y)/class3n + 1e-8)))
        gain = en - class1n/total_n * en1 - class2n/total_n * en2 - class3n/total_n * en3

    print(f"en  = {en:0.4}")
    print(f"en1 = {en1:0.4}")
    print(f"en2 = {en2:0.4}")
    print(f"en3 = {en3:0.4f}")
    print(f"gain = {gain}")

cal_en(total_n, class1n, class2n, class3n, class1y, class2y, class3y)
```

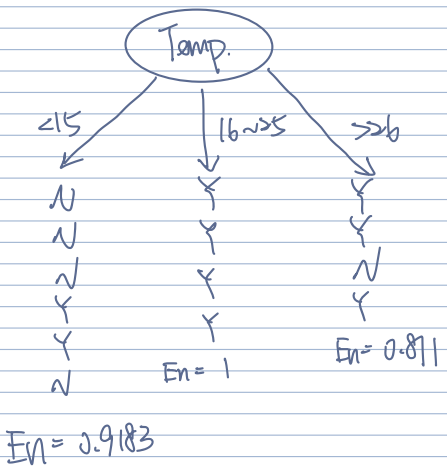
✓ 0.4s

```
en  = 0.9852
en1 = 0.9183
en2 = -1.443e-08
en3 = -0.0000
gain = 0.5916727703833556
```

利用設計的程式計算各層的information gain

1. first layer.

(a) Temp



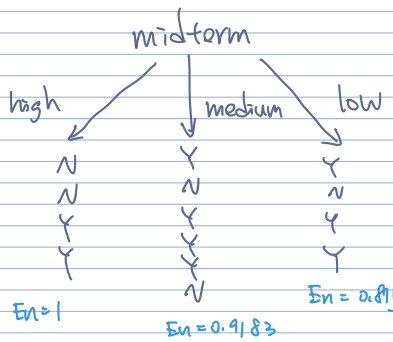
$$En(t) = -\left(\frac{9}{14}\log_2\left(\frac{9}{14}\right) + \frac{5}{14}\log_2\left(\frac{5}{14}\right)\right)$$

$$= 0.9403$$

$$Gain = 0.9403 - \frac{6}{14}\times 0.9183 - \frac{4}{14}\times 1 - \frac{4}{14}\times 0.811$$

$$= 0.0293$$

(b) midterm



En(t) = 0.9403

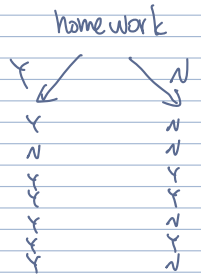
En(t1) = 0.9403

En(t2) = 1

En(t3) = 0.8113

gain = 0.0292

(c) homework



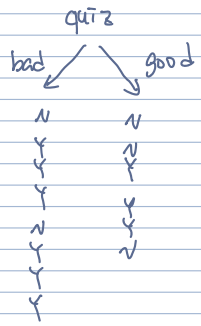
En(t) = 0.9403

En(t1) = 0.5917

En(t2) = 0.9852

gain = 0.1518

(d) quiz



En(t) = 0.9403

En(t1) = 0.891

En(t2) = 1

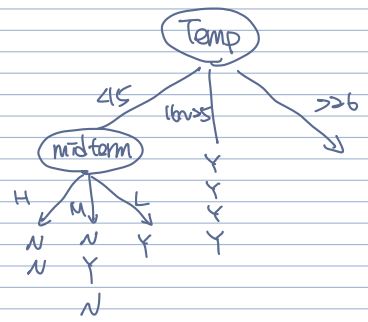
gain = 0.04813

Temp 的 information gain 最大，雙贏做為 first layer.

2. second layer

(a)

① <15, mid



En(k) = 0.9183

En(k1) = 0

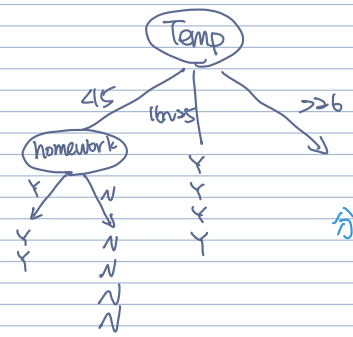
En(k2) = 0.9183

En(k3) = 0

Gain = 0.459

(a)

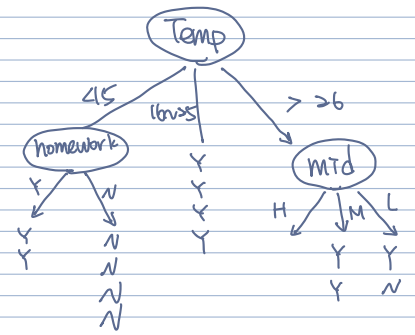
② <15, homework



分出来了, 要 homework

(b) >>6, mid

①



En(k) = 0.813

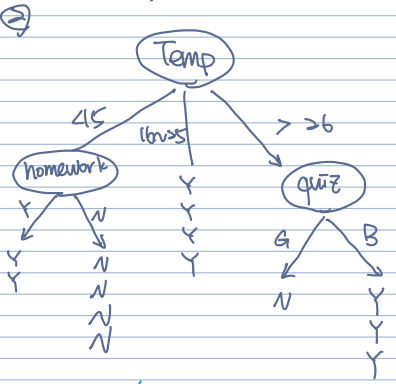
En(k1) = 1

En(k2) = 0

gain = 0.3112

(b) mid, quiz

②



分出来了, 要 gain

⇒ Final result.