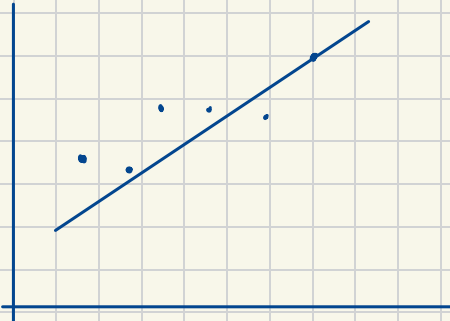


ML + AI



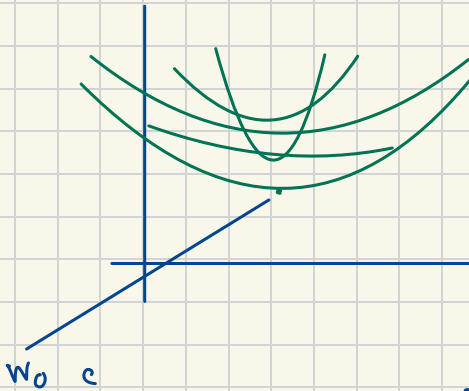
$$y = mx + c$$

State =

m	c
---	---

0.38	0.46
------	------

GA



Gradient Descent

หาค่าความชันที่ต่ำสุดจริง ๆ

100 จุด
N. $m \sim [1.0, 2.0]$ 0.01

$c \sim [1.0, 2.0]$ 0.01
100 จุด

จำนวน Search = 10,000 จุด

Classification

	Sky	Temp	Humid	Wind	Water	Forecast	EnjoySpt
D1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
D2	Sunny	Warm	High	Strong	Warm	Same	Yes
D3	Rainy	Cold	High	Strong	Warm	Change	No
D4	Sunny	Warm	High	Strong	Cool	Change	Yes

D5 Sunny cold Normal Strong Warm Change ?

① Hypothesis Representation $h(\vec{x})$

$h(< a_1, a_2, a_3, a_4, a_5, a_6 >):$
 $h(< \text{Sunny}, \text{Warm}, \text{Normal}, \text{Strong}, \text{Warm}, \text{Same} >):$

if $a_1 == \text{'Sunny'}$ and $a_5 == \text{'Warm'}$:

return Yes

return No

unknown state

Sunny	?	?	?	Warm	?
-------	---	---	---	------	---

unknown state

naive hypothesis search

Rainy Warm Normal Strong Cool Same
 ? cold High Normal ? Change

m Combination of match

Yes

3	3	3	3	3	3
Sunny	?	?	?	?	?

?	Warm	?	?	?	?
---	------	---	---	---	---

Sunny	Warm	?	?	?	?
-------	------	---	---	---	---

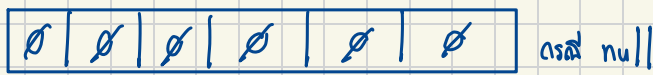
3^6 (any) +
 1 ~~1 1 1 1 1 1~~
 3^{6+1} - Hypothesis space

naive search

naive search method

a_1	a_2	a_3	a_4	a_5	a_6	Ans
_____	_____	_____	_____	_____	_____	No
_____	_____	_____	_____	_____	_____	No
_____	_____	_____	_____	_____	_____	No
_____	_____	_____	_____	_____	_____	No

Brute Force



ถ้ามีตัวอย่างที่มี representation ที่ จะสามารถให้ข้อโต้แย้ง

Maintain hypothesis ที่ Specific ที่สอดคล้องกับข้อมูลที่พบมาทั้งหมด

so $\langle \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset \rangle$ มีลักษณะที่ negative
ไม่ครอบคลุมโดยเลย (Null hypothesis)

D1 $\langle \text{Sunny, Norm, Normal, Strong, Warm, Same} \rangle \rightarrow$ ทำให้ so general ครอบคลุม

แล้ว S1 $\langle \text{Sunny, Norm, } \boxed{\text{Normal}}, \text{Strong, Warm, } \boxed{\text{Same}} \rangle$ เป็นตัวอย่าง D1

D2 $\langle \text{Sunny, Norm, } \boxed{\text{High}}, \text{Strong, Warm, } \boxed{\text{Change}} \rangle$

แล้ว S1 ยังไม่ครอบคลุม D2

S3, S2 $\langle \text{Sunny, Norm, } \textcircled{?}, \text{Strong, Warm, } \textcircled{?} \rangle$ เพิ่มอะไร? ไม่ครอบคลุม Normal, High ลบออก 3 ที่ขึ้น Negative (No)

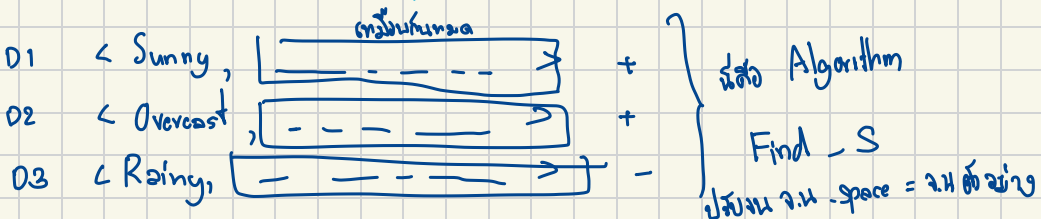
D4 $\langle \text{Sunny, Warm, High, Strong, Cool, Change} \rangle$

S4 $\langle \text{Sunny, Norm, } ?, \text{Strong, } ?, ? \rangle \rightarrow$ ครอบคลุม D1, 2, 4

ทำไมจึงไม่ Negative ได้

① เรา generalize มนุษย์ที่คิดว่าที่เห็นไม่พอ และ

② มีตัวอย่าง (ซึ่งนับกับว่าตัวอย่างใน representation มนุษย์)



S1 < Sunny, [] >

S2 < ? , [] >

มันจะไปรวมกับ P3 ที่มัน - มันก็
จะทำมันให้มันอยู่ใน representation
แล้วใน representation อันนี้

Sunny or Overcast

→ Rainy

อันนี้ V กับ อันที่มันรวม
กับ มัน อันที่ space มันมี

Sunny V Rainy

- Sunny
- Rainy
- Overcast

** Find_S จะไม่ได้คือของมีค่ารวม (มันบอกว่ามีค่ารวม representation มัน)

อันนี้ให้เรียกว่า Inductive Bias ที่มันรวมค่าของมันที่ใส่เข้ามา

Turn Linear Regression, Enjoy Sport

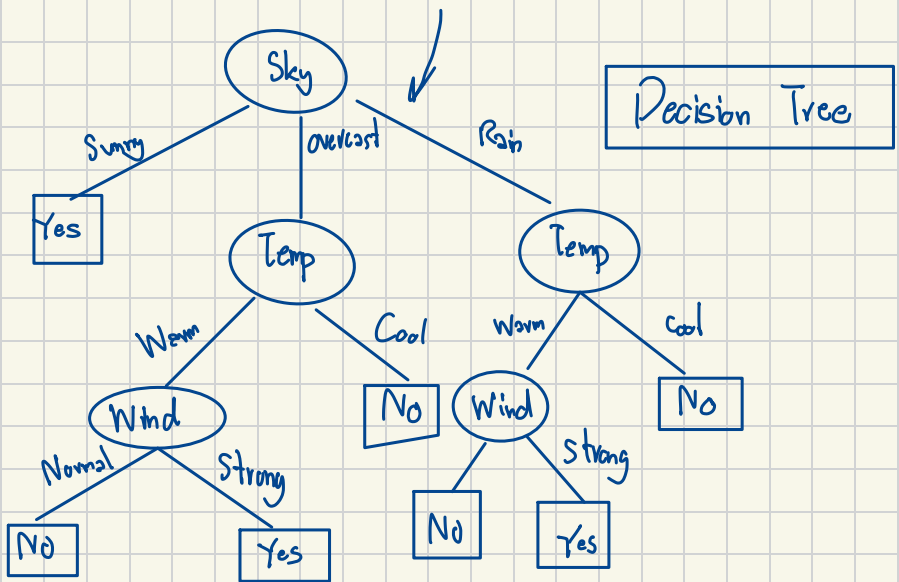
Find - S

2 2 2
< 21, 22, 23, 24, 25, 26 >

< Sunny, ? , ? , ? , ? , ? >

< ? , Norm, ? , Strong, ? , ? >

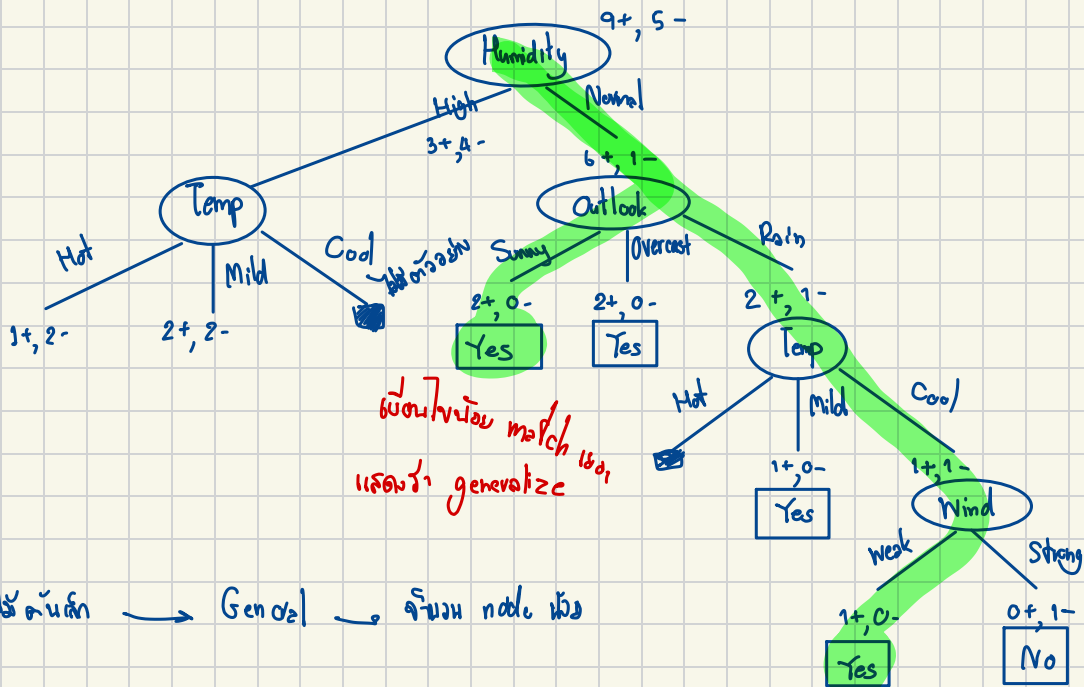




พยายามอธิบายให้เข้าใจด้วย ต้นไม้

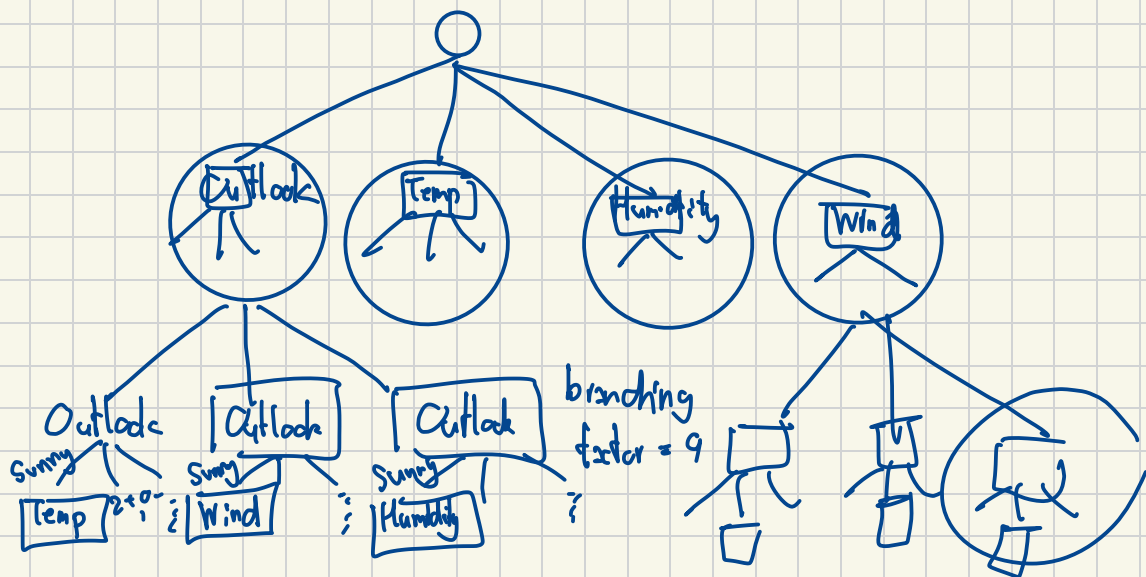
Yes 9 ครั้ง No 5 ครั้ง

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



အဲဒါကမူလ → General → အဲဒါက match လာ

အဲဒါက match လာ
အဲဒါက specific

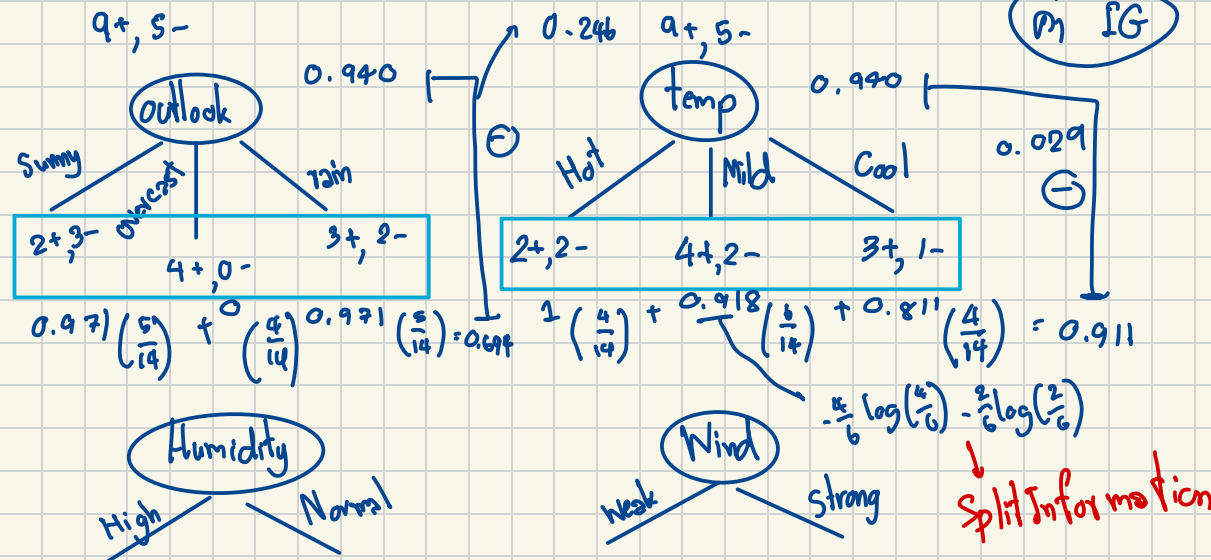
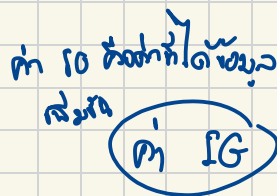
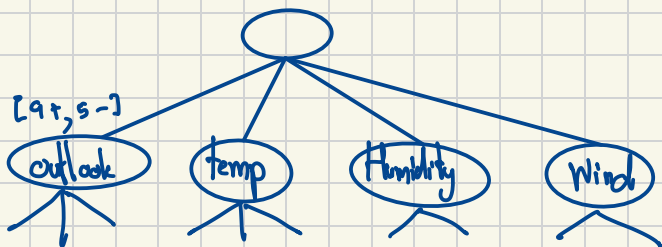


for BFS \Rightarrow Optimal \rightarrow optimal solution

Greedy ไม่รับประกันว่าได้ Prefer ดีกว่า
ถ้าไม่พบคำตอบที่ดีที่สุด ทิ้งไว้

Greedy 11/11/2023

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

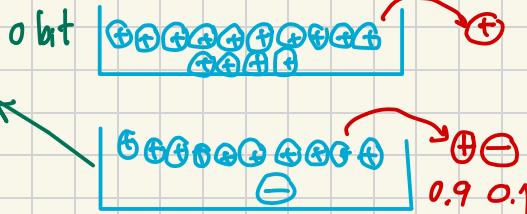


Entropy | Information Theory

$$\text{Entropy (S)} = \sum_v -p_v \log_2 p_v$$



$$E_{\text{entropy}}([9, 1]) = -\frac{9}{10} \log_2 \frac{9}{10} - \frac{1}{10} \log_2 \frac{1}{10} = 0.4689 \text{ bit}$$



จำนวน bit ของ information ที่สามารถ

ใช้แทนเหตุการณ์ที่ได้อีก

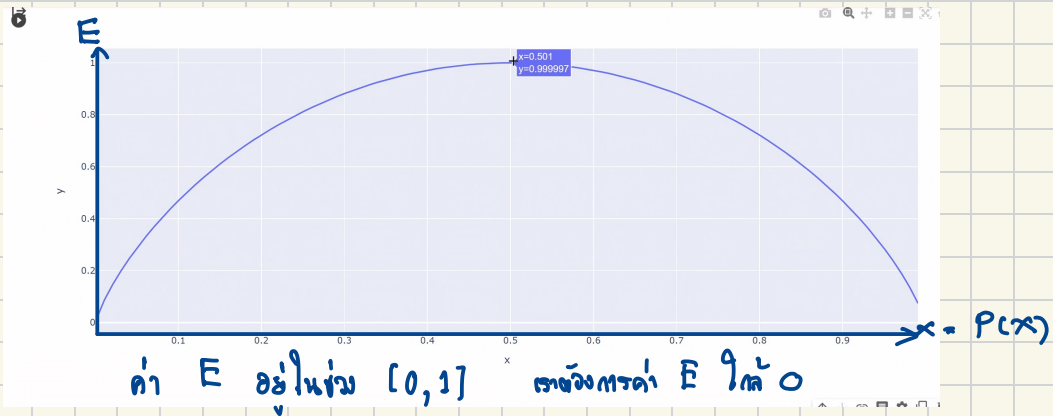


$\oplus \ominus$
0.5 0.5

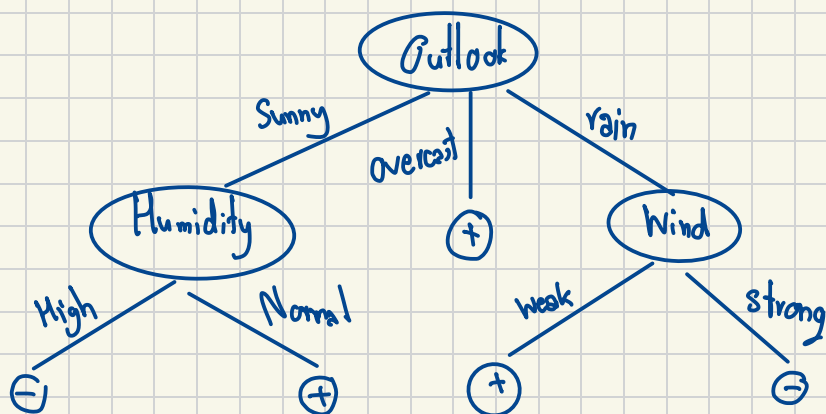
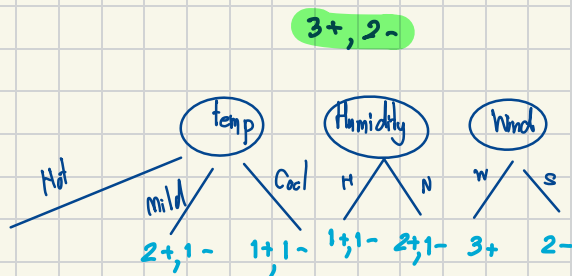
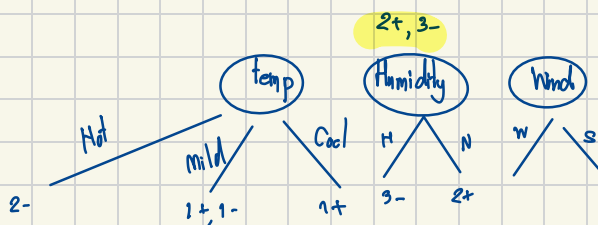
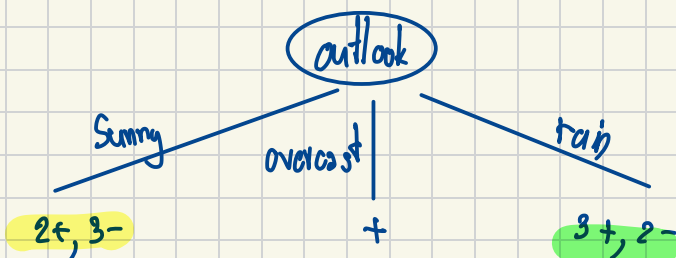
$$\text{Entropy} [20+, 0-] = 0 \text{ bit}$$

$$\text{Entropy} [10+, 10-] = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2} = 1 \text{ bit}$$

สิ่งนี้จะเป็น bit คือใช้แทนเหตุการณ์

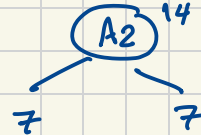
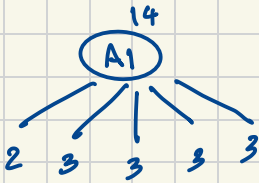


Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No



$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Information}(S, A)}$$

$$\text{Split Information}(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$



ถ้า Gain 2 ต่ำกว่า Gain Ratio

A1

$$\text{Split Inform} = -\frac{2}{14} \log_2 \frac{2}{14} - \frac{3}{14} \log_2 \frac{3}{14} - \frac{3}{14} \log_2 \frac{3}{14} - \frac{3}{14} \log_2 \frac{3}{14} - \frac{3}{14} \log_2 \frac{3}{14}$$

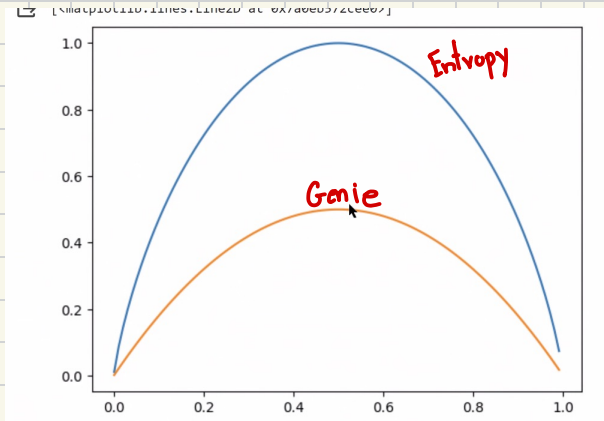
$$= 2.306$$

A2

$$\text{Split Inform} = -\frac{7}{14} \log_2 \frac{7}{14} - \frac{7}{14} \log_2 \frac{7}{14} = 1 \text{ คุ้มกว่า}$$

Genic impurity

$$\sum_p^J p_i (1-p_i)$$



Genic ดีกว่า [0, 0.5]

ถ้า default von sklearn