BIAS

- Bias: If your model has high bias it means that it has high training error rate and if it has low Bias it mean it has low traing rate
- Algorithms which have high Bias learn faster and are easier to understand but generally less flexible. In turn they have lower predictive performance on complex problems that fail to meet the simplifying assumption of the algorithm bias.

VARIANCE

- Variance - If your model has high variance it means that it has high testing error rate and if it has low Variance it means it has low testing error rate.
- ML Algos that have a high variance are strongly influenced by the specifics of the training data. This means that the specifics of the training have influences the number and types of parameters used to characterize the mapping function.

TRADE OFF

- Increasing the bias will decrease the variance and vice versa. There is a <u>trade-off</u> at play between these two concerns and the algorithms you choose and the way you choose to configure them are finding different balances in this trade-off

	Class 1	Class 2
Class 1	50	30
Class 2	40	60

1) Precision:
Assuming Class 1 to be positive & Class 2 to be negative

$P_1 \Rightarrow \frac{TP}{TP+FP} \qquad \frac{50}{80}$

$P_2 \rightarrow$ Assuming Class 2 to be positive and class 2 to be negative

$\Rightarrow \frac{TP}{TP+FP} = \frac{40}{100}$

2) Recall $= \frac{TP}{TP+FN}$

a) When Class 1 is positive
$\Rightarrow \frac{50}{90} = \frac{5}{9}$

b) When Class 2 is posit
$= \frac{40}{90}$

→ To reduce bias → you can increase the model complexity and have more parameters to tune & allow for more learning.

→ Reducing variance → add more data & increase regularization to prevent model from overfitting → Reduce dimensionality by removing redundant data.

→

3)

3) F1 score $= 2 \cdot \dfrac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$

$$= \frac{2 \cdot 50/90 \cdot 50/80}{\frac{50}{90} + \frac{50}{80}}$$

$$\Rightarrow \frac{2 \times \frac{5}{9} * \frac{5}{8}}{\frac{85}{72}}$$

$$\Rightarrow \frac{25}{36} \div \frac{85}{72}$$

$$\Rightarrow \frac{25}{\cancel{36}} \times \frac{\overset{2}{\cancel{72}}}{85} \Rightarrow \frac{5}{9}$$

Gain (Wind) = 0.81

W	Y	N
W	3	0
S	0	1

highest gain = wind

Taking sunny side

O	T	H	W	P
S	H	H	W	N
S	H	H	S	N
S	M	H	W	N
S	C	N	W	Y

$E_{sunny} = -\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log 3/4 = 0.81$

$E_{weak} = 0.918$

Wind	Y	N
W	1	2
S	0	1

$E_{strong} = 0$

$\text{Gain} = 0.811 - \left(\frac{3}{4} \times 0.918\right) = 0.1225$

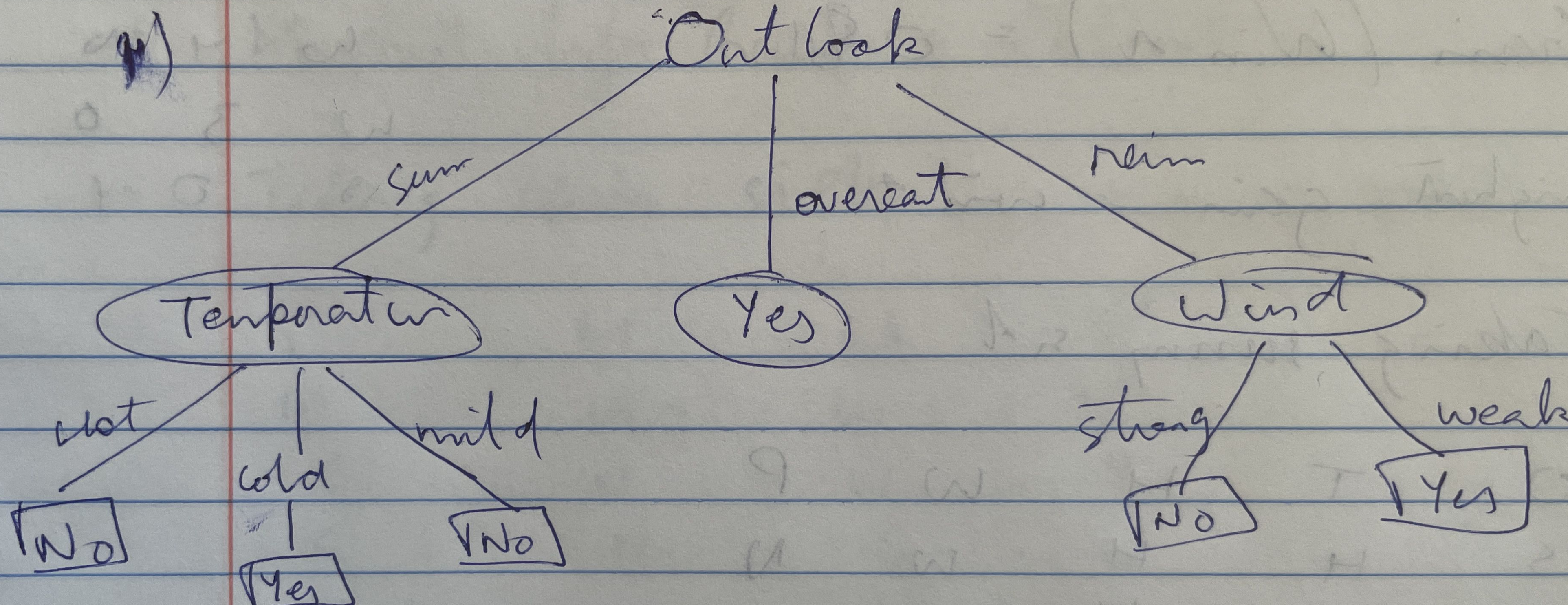
Gain temp = 0.811

T	Y	N	E
H	0	2	0
M	0	1	0
C	1	0	0

Gain humidity = 0.811

H	Y	N
high	0	3
normal	1	0

4)



Out look
sunn
overcast
rain
Temperatur
Yes
Wind
Hot
cold
mild
No
Yes
No
strong
weak
No
Yes

$E_h = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.91$

	Y	N
H	1	2
M	2	1
C	3	1

$E_M = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.91$

$E_C = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.81$

Gain (temp) $= 0.971 - (0.3 \times 0.91) - (0.3 \times 0.91) - (0.4 \times 0.81)$
$= 0.096$

$E_h = -\frac{2}{5}\log\frac{2}{5} - \frac{3}{5}\log\frac{3}{5} = 0.971$

H	Y	N
H	2	3
N	4	1

$E_N = -\frac{4}{5}\log\frac{4}{5} - \frac{1}{5}\log\frac{1}{5} = 0.72$

Gain (humidity) $= 0.971 - (0.5 \times 0.97) - (0.5 \times 0.72)$
$= 0.125$

$E_S = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} = 0.91$

W	Y	N
S	1	2
W	5	2

$E_W = -\frac{5}{7}\log\frac{5}{7} - \frac{2}{7}\log\frac{2}{7} = 0.86$

Gain (wind) $= 0.971 - (0.3 \times 0.91) - (0.7 \times 0.86)$

$= 0.091$

3) $Entropy = \frac{-p}{p+n} \log \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$

$\Rightarrow -\frac{6}{10} \log \frac{6}{10} - \frac{4}{10} \log \frac{4}{10} = 0.971$

O	Y	N
Sunny	1	3
Overcast	2	0
Rain	3	1

$\rightarrow E_{sunny} = \frac{-1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0.811$

$E_{overcast} = -\frac{2}{2} \log \frac{2}{2} = 0$

$E_{rain} = -\frac{3}{4} \log \frac{3}{4} - \frac{1}{4} \log \frac{1}{4} = 0.811$

$Gain(outlook) = 0.971 - \left(\frac{4}{10} \cdot 0.8111\right) - (0.2 \cdot 0) - (0.4 \times 0.81)$

$= 0.32$

→ Highest Info Gain = Outlook

→ Taking rain side

T	H	W	P
M	H	W	Y
C	N	W	Y
C	N	S	N
M	N	W	Y

→ $E_{rain} = -\frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0.81$

$E_m = 0$

$E_c = 1$

T	Y	N
M	2	0
C	1	1

Gain temp : $0.81 - \left(\frac{2}{4} \times 1\right) = 0.311$

$E_h = 0$

$E_n = -\frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3}$

$= 0.918$

H	Y	N
H	1	0
N	2	1

Gain = $0.811 - \left(\frac{3}{4} \times 0.918\right)$

$= 0.1225$