ML FALL 2018 | HOMEWORK 3 | PART I | amc1354 & ads798

PROBLEM #1

S: $N^+ = 4$, $N^- = 5$, $N = N^+ + N^- = 9$

S': $N^+ = 3$, $N^- = 6$, $N = 9$

We expect S to have higher entropy than S' because entropy describes the rate of information in a sample. In other words, it relates to the balance of a sample. S is more balanced than S'. Therefore, it should have higher entropy. In fact, without using a calculator:

$$\text{Entropy}(S) = -\left[\frac{N^+}{N}\log_2\frac{N^+}{N} + \frac{N^-}{N}\log_2\frac{N^-}{N}\right]$$

$$= -\left[\frac{N^+}{N}\left(\log_2 N^+ - \log_2 N\right) + \frac{N^-}{N}\left(\log_2 N^- - \log_2 N\right)\right]$$

$$= -\frac{1}{N}\left[N^+\log_2 N^+ - N^+\log_2 N + N^-\log_2 N^- - N^-\log_2 N\right]$$

$$= -\frac{1}{N}\left[N^+\log_2 N^+ + N^-\log_2 N^- - (N^+ + N^-)\log_2 N\right]$$

$$= -\frac{1}{N}\left(N^+\log_2 N^+ + N^-\log_2 N^- - N\log_2 N\right)$$

For S and S', N is the same, namely 9. So the term to compare is actually $N^+\log N^+ + N^-\log N^-$. In fact,

$$\text{Entropy}(S) \overset{?}{\lessgtr} \text{Entropy}(S')$$

$$-\frac{1}{\cancel{9}}\left(4\log_2 4 + 5\log_2 5 - \cancel{9\log_2 9}\right) \lessgtr -\frac{1}{\cancel{9}}\left(3\log_2 3 + 6\log_2 6 - \cancel{9\log_2 9}\right)$$

$$-\left(4\log_2 4 + 5\log_2 5\right) \lesseqgtr -\left(3\log_2 3 + 6\log_2 6\right)$$

$$\log_2\left(4^4 \cdot 5^5\right)^{-1} \lesseqgtr \log_2\left(3^3 \cdot 6^6\right)^{-1}$$

$$\frac{1}{4^4 \cdot 5^5} \lesseqgtr \frac{1}{3^3 \cdot 6^6}$$

$$\frac{3^3 \cdot 2^6 \cdot 3^6}{2^4 \cdot 2^4 \cdot 5^5} \lesseqgtr 1$$

$$\frac{3^9}{2^2 \cdot 5^5} > \frac{3^9}{2^2 \cdot 3^5} = \frac{3^4}{2^2} = \frac{81}{4} > 1$$

$\Rightarrow$ Entropy $(S) >$ Entropy (S')

(b) In x_1, we have: $V = \{F, T\}$, $S_F = \{+, -, +, -\}$, $S_T = \{+, -, -\}$
$S = \{+, +, -, +, -, -, -\}$

$$x_1 = \begin{bmatrix} F \\ T \\ F \\ F \\ T \\ T \\ F \end{bmatrix}$$

$$\text{Entropy}(S_F) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$$

$$\text{Entropy}(S_T) = -\frac{1}{3}\log_2\frac{1}{3} - \frac{2}{3}\log_2\frac{2}{3} = 0.918$$

$$\text{Entropy}(S) = -\frac{3}{7}\log_2\frac{3}{7} - \frac{4}{7}\log_2\frac{4}{7} = 0.985$$

$$IG_1(S) = \text{Entropy}(S) - \frac{4}{7}\text{Entropy}(S_F) - \frac{3}{7}\text{Entropy}(S_T)$$

$$= 0.985 - \frac{4}{7} \cdot 1 - \frac{3}{7} \cdot 0.918 = 0.0202$$

(c)

i	x_1	x_2	x_3	r
1	F	F	F	+
2	T	F	T	+
3	F	F	T	−
4	F	T	F	+
5	T	T	F	−
6	T	T	T	−
7	F	F	F	−

$IG_1 = 0.02$

$IG_2 = 0.02$

$IG_3 = 0.02$

WE SPLIT ON x_1
(WE PICK RANDOMLY AS THEY ARE ALL EQUAL)

$$IG_2 = 0.985 - \frac{4}{7} \cdot 1 - \frac{3}{7} \cdot 0.918 = 0.02$$

$S_F = \{+ + - -\}$

$S_T = \{+ - -\}$

$IG_3 = \text{SAME}$

$S_F = \{+ + - -\}$

$S_T = \{+ - -\}$

FIRST PART OF TREE:

$[x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}]$

$x_1 = T$: $[x^{(2)}, x^{(5)}, x^{(6)}]$ *

$x_1 = F$: $[x^{(1)}, x^{(3)}, x^{(4)}, x^{(7)}]$ **

i	x_2	x_3	r
2	F	T	+
5	T	F	−
6	T	T	−

$$IG_2 = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} - \frac{1}{3}(0) - \frac{2}{3}(0) = 0.918$$

$S_F = \{+\}$

$S_T = \{- \, -\}$

$S = \{+ \, - \, -\}$

$$IG_3 = -\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\cdot(0) - \frac{2}{3}\cdot(1) = 0.25$$

$S_F = \{-\}$

$S_T = \{+, -\}$

$S = \{+ \, - \, -\}$

$\Rightarrow$ we split on x_2 because for x_2, the IG is higher than the IG for x_3.

i	x_2	x_3	r
1	F	F	+
3	F	T	−
4	T	F	+
7	F	F	−

$$IG_2 = 1 - \frac{1}{4}(0) - \frac{3}{4}(0.918) = 0.31$$

$S_T = \{+\}$

$S_F = \{+ \, - \, -\}$

$S = \{+ \, + \, - \, -\}$

SAME

$$IG_3 = 1 - \frac{1}{4}(0) - \frac{3}{4}(0.918) = 0.31$$

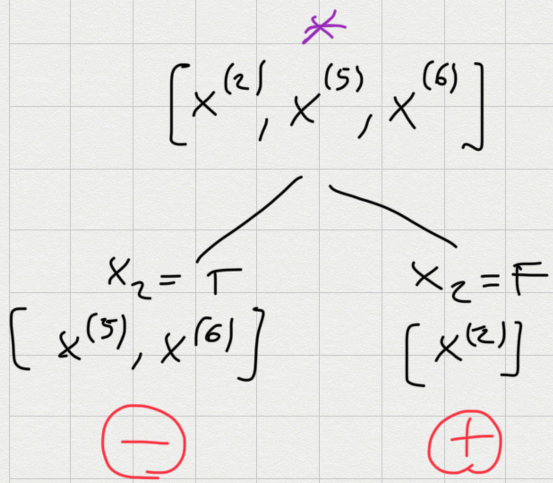
$S_T = \{-\}$

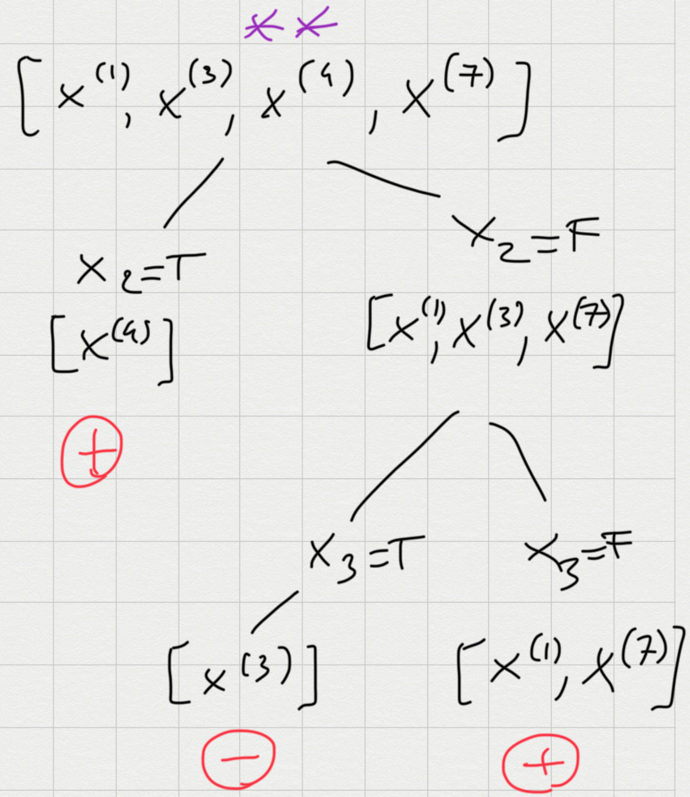
$S_F = \{+ \, + \, -\}$

$S = \{+ \, + \, - \, -\}$

$\Rightarrow$ WE PICK RANDOMLY x_2

SECOND PART OF TREE:





Here x_3 has different values but the label is − for both $x_3 = F$ and $x_3 = T$

Here it is + because there is a tie

Training accuracy can't be 100%

(d) $$H(Y) = -\sum_{i=1}^{n} P(Y=i) \log_2 P(Y=i)$$

$$H(Y|X_1) = \sum_{x} P[X_1 = x] \cdot \left(\sum_{y} - P[Y=y \mid X=x] \log_2 P[Y=y \mid X=x] \right)$$

$H(Y) - H(Y|X)$ mutual information $= H(X) - H(X|Y)$

$\hookrightarrow IG(X_1)$

$$H(Y) = -\left[\mathbb{P}(Y=+)\log_2 \mathbb{P}(Y=+) + \mathbb{P}(Y=-)\log_2 \mathbb{P}(Y=-)\right]$$

$$= -\left[\frac{3}{7}\log_2 \frac{3}{7} + \frac{4}{7}\log_2 \frac{4}{7}\right] = 0.985$$

$$H(Y|X_1) = \mathbb{P}(X_1=T)\left[-\mathbb{P}(Y=+|X_1=T)\log_2 \mathbb{P}(Y=+|X_1=T) + \right.$$

$$\left. - \mathbb{P}(Y=-|X_1=T)\log_2 \mathbb{P}(Y=-|X_1=T)\right] +$$

$$+\mathbb{P}(X_1=F)\left[-\mathbb{P}(Y=+|X_1=F)\log_2 \mathbb{P}(Y=+|X_1=F) + \right.$$

$$\left. - \mathbb{P}(Y=-|X_1=F)\log_2 \mathbb{P}(Y=-|X_1=F)\right]$$

$$= \frac{3}{7} \cdot \left[-\frac{1}{3}\log_2 \frac{1}{3} - \frac{2}{3}\log_2 \frac{2}{3}\right] +$$

$$+ \frac{4}{7}\left[-\frac{2}{4}\log_2 \frac{2}{4} - \frac{2}{4}\log_2 \frac{2}{4}\right] = \frac{3}{7} \cdot 0.918 + \frac{4}{7} \cdot 1$$

$$= 0.96486$$

$$H(Y) - H(Y|X_1) = 0.985 - 0.965 = \underline{0.0202}$$

(e) $S = \{ \underbrace{l_1, l_1, \dots, l_1}_{\frac{N}{z}}, \underbrace{l_2, \dots,}_{\frac{N}{z}} \underbrace{l_3}_{\frac{N}{z}}, \underbrace{\dots, l_z}_{\frac{N}{z}} \}$ $\#S = N$

GROUPS OF EQUAL LABELS

$$L = \{ l_1, l_2, \dots, l_z \} \quad \# L = z$$

$$\text{Entropy}(S) = - \sum_{l \in L} \frac{N_l}{N} \log_2 \frac{N_l}{N}$$

$$= - \sum_{l \in L} \frac{\frac{\cancel{N}}{z}}{\cancel{N}} \log_2 \frac{\frac{\cancel{N}}{z}}{\cancel{N}}$$

$$= - \sum_{l \in L} \frac{1}{z} \log \frac{1}{z}$$

$$= - \cancel{z} \cdot \frac{1}{\cancel{z}} \log \frac{1}{z} = \underline{\log z}$$

ML FALL 2018 | HOMEWORK 3 | PART II / "calculate by hand questions"

amc1354 & ads798

PROBLEM # 2

(a) $A - \lambda I = \begin{bmatrix} 0 & 14 \\ 6 & 9 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$= \begin{bmatrix} -\lambda & 14 \\ 6 & 9-\lambda \end{bmatrix}$

$|A - \lambda I| = -\lambda(9-\lambda) - 14 \cdot 6 = -9\lambda + \lambda^2 - 84$

$= \lambda^2 - 9\lambda - 84$

(b) $\lambda^2 - 9\lambda - 84 = 0$

$\lambda_{1,2} = \dfrac{9 \pm \sqrt{81 + 4 \cdot 84}}{2} = \dfrac{9}{2} \pm \dfrac{\sqrt{417}}{2}$ $\begin{cases} \underline{\underline{14.71}} \\ \underline{\underline{-5.71}} \end{cases}$

(c) $\lambda_1 = 14.71$

$A v_1 = \lambda_1 v_1 \quad \begin{bmatrix} -14.71 & 14 \\ 6 & 9-14.71 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$-14.71 \cdot x_1 + 14\, y_1 = 0$$

$x_1 = \frac{14}{14.71} y_1$ so, eigenvectors are of the form $\begin{bmatrix} 1 \\ 14/14.71 \end{bmatrix} x_1$

Now, adding the condition $\sqrt{x_1^2 + y_1^2} = 1$, we have:

$$\sqrt{1^2 + \left(\frac{14}{14.71}\right)^2}\; x_1 = 1 \Rightarrow x_1 = \pm 0.7244 \Rightarrow v_1 = \begin{bmatrix} \pm 0.7244 \\ \pm 0.6894 \end{bmatrix}$$

$$y_1 = \pm 0.7244 \cdot \frac{14}{14.71}$$

$$\lambda_2 = -5.71029$$

$$\begin{bmatrix} +5.71 & 14 \\ 6 & 9+5.71 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$5.71\, x_2 + 14\, y_2 = 0 \Rightarrow \begin{cases} x_2 = -\frac{14}{5.71} y_2 \\ \sqrt{x_2^2 + y_2^2} = 1 \end{cases}$$

$$\sqrt{1^2 + \left(\frac{14}{5.71}\right)^2}\; x_2 = 1 \Rightarrow x_2 = \pm 0.3777 \Rightarrow v_2 = \begin{bmatrix} \pm 0.3777 \\ \mp 0.9259 \end{bmatrix}$$

$$\Rightarrow y_2 = -(\pm 0.378) \cdot \frac{14}{5.71}$$

(d) import numpy as np

```
A = np.matrix([[0, 14],[6, 9]])

[V, D] = np.eig(A)

linalg.eig(A)
```

... CHECK OTHER PDF FILE

PROBLEM #3

(a)

x_1	x_2	x_3
5	2	4
9	6	4
7	1	0
2	5	6
mean: 5.75	3.5	3.5

centering $\longrightarrow$

x_1^c	x_2^c	x_3^c
−0.75	−1.5	0.5
3.25	2.5	0.5
1.25	−2.5	−3.5
−3.75	1.5	2.5

$$\text{mean}(x_1) = \frac{5+9+7+2}{4} = \frac{23}{4} = 5.75 \qquad \text{mean}(x_2) = \frac{2+6+1+5}{4} = \frac{14}{4} = 3.5$$

$$\text{mean}(x_3) = \frac{4+4+6}{4} = 2 + \frac{3}{2} = 3.5$$

$$B = \begin{bmatrix} -0.75 & -1.5 & 0.5 \\ 3.25 & 2.5 & 0.5 \\ 1.25 & -2.5 & -3.5 \\ -3.75 & 1.5 & 2.5 \end{bmatrix}$$

(b) $s_{1,3} = \dfrac{\sum_t (x_1^t - s_1)(x_3^t - s_3)}{N-1}$ covariance in B

note that $s_1^c = s_2^c = s_3^c = 0$ in B

$$s_{1,3} = \frac{-0.75 \cdot 0.5 + 3.25 \cdot 0.5 + 1.25 \cdot (-3.5) - 3.75 \cdot 2.5}{4-1}$$

$= -4.16667$

(c) B = np.matrix[...]

np.cov(B) ← check entry $\boxed{1,3}$

max(linalg.eig(B)) ... CHECK OTHER PDF FILE

(d) Run code and output 1st 2 columns of Z

... CHECK OTHER PDF FILE

$\boxed{\text{PROBLEM \# 5}}$ RUN ENTIRELY ON PYTHON ... CHECK OTHER PDF FILE

(#4, ACTUALLY)