CSE 512: Final Exam Dec 9,10

Instructions:

- You are encouraged to use no more than 3 hours to complete the exam, but you may use as long as you wish within the 48 hour window.
- You are allowed 1 page (front and back) cheat sheet. The cheat sheet must be scanned (or photographed with high resolution) and submitted along with the exam solutions. The time begins when you flip this page.
- You are also allowed a simple calculator, or Wolfram Mathematica to simplify equations. You may also write simple scripts in python or MATLAB to aid with your calculation. However, you do not need to for solving the problem, and you do not need to submit any code.
- You may print the exam and write your solutions or use lyx/latex. If you need extra sheets of paper, please label them carefully as to which question they are answering. Make your final answer clear.
- If you choose to handwrite your solutions, you must make sure that the digital scan / photograph is of high enough quality that we can see everything clearly. Anything we can't read, we will not grade.
- You may not discuss any problem with any other student while the exam submission portal is still open. You may not look for answers on the internet or in any notes outside of your cheatsheet.

Name: Venkata Subba Narasa Bharath Meadam

Student ID: 112672986

Scoring	
Q 1	________ / 10
Q 2	________ / 10
Q 3	________ / 20
Q 4	________ / 20
Q 5	________ / 20
Q 6	________ / 20
Total	________ / 100

1. Classify the following as a supervised learning task (where training labels are needed) or unsupervised learning task (where training labels are not needed) **(2 point each)**

   (a) Principal component analysis

   Supervised **Unsupervised**

   (b) logistic regression

   Supervised **Unsupervised**

   (c) Gaussian mixture model

   Supervised **Unsupervised**

   (d) boosted decision trees

   Supervised **Unsupervised**

   (e) clustering

   Supervised **Unsupervised**

2. **True or False. (2 point each)**

   (a) A graphical model with on average 2 directed edges pointing to each node requires less training samples for inference, than a similar graphical model with on average 5 directed edges pointing to each node.

   True **False**

   (b) In a hidden Markov model, each hidden state is independent from all other hidden states

   True **False**

   (c) In multiclass logistic regression, the event that a sample x belongs to class i is independent of the event that x belongs to class j whenever $i \neq j$.

   True **False**

   (d) The difference between K-means and K-NN is that K-means is a supervised learning task and K-NN is an unsupervised learning task.

   True **False**

   (e) Ensembling different models presents more advantages if each model acts as independently as possible.

   True **False**

Q3)

a) entropy :- $H(X) = \sum_{X=x} Pr(X=x) \log_2 (Pr(X=x))$

$Pr(\text{label} = \text{aardvark}) = \frac{2}{8} = \frac{1}{4}$

$Pr(\text{label} = \text{rabit}) = \frac{3}{8}$

$Pr(\text{label} = \text{monkey}) = \frac{2}{8} = \frac{1}{4}$

$Pr(\text{label} = \text{moose}) = \frac{1}{8}$

$$H(X) = -\left(\frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{3}{8}\log_2\left(\frac{3}{8}\right) + \frac{1}{4}\log_2\left(\frac{1}{4}\right) + \frac{1}{8}\log_2\left(\frac{1}{8}\right)\right)$$

$$= 1.906$$

$$\boxed{\therefore \text{Entropy} = 1.906}$$

(b) Information gain

(i) Nose :-

$$H(X|Y) = -\sum_{X=x, Y=y} Pr(X=x, Y=y) \log_2 Pr(X=x|Y=y)$$

$$H(\text{nose}=\text{small}) = -\left(\frac{2}{5}\log\left(\frac{2}{5}\right) + \frac{3}{5}\log\left(\frac{3}{5}\right) + 0 + 0\right)$$

$$= 0.971$$

$$H(\text{nose}=\text{big}) = -\left(0 + 0 + \frac{2}{3}\log\left(\frac{2}{3}\right) + \frac{1}{3}\log\left(\frac{1}{3}\right)\right)$$

$$= 0.918$$

$I(\text{nose}) = P(\text{small}) * H(\text{nose} = \text{small} \mid Y)$
$\quad + P(\text{big}) * H(\text{nose} = \text{big} \mid Y)$

$= \frac{5}{8} \times 0.971 + \frac{3}{8} \times 0.918$

$= 0.951$

Information Gain $\Rightarrow H(S) - I(\text{Nose}) = 1.906 - 0.951$
$= \boxed{0.95}$

(ii) Ear Shape :-

$H(\text{ear shape} = \text{round}) = -\left(\frac{2}{4}\log_2\left(\frac{2}{4}\right) + 0 + \frac{2}{4}\log_2\left(\frac{2}{4}\right) + 0\right)$
$= 1$

$H(\text{ear shape} = \text{long}) = -\left(0 + \frac{3}{4}\log\left(\frac{3}{4}\right) + 0 + \frac{1}{4}\log\left(\frac{1}{4}\right)\right)$

$= 0.811$

$I(\text{ear shape}) = \frac{4}{8} \times 1 + \frac{4}{8} \times 0.811$

$= 0.906$

Information Gain $\Rightarrow H(S) - I(\text{ear shape}) = 1.906 - 0.906$
$= \boxed{1.0}$

(iii)
Ear Position :-

$H(\text{ear Position} = \text{High}) = -\left(\frac{2}{6}\log\left(\frac{2}{6}\right) + \frac{3}{6}\log\left(\frac{3}{6}\right) + \frac{1}{6}\log\left(\frac{1}{6}\right) + 0\right)$

$= 1.459$

$H(\text{ear Position} = \text{low}) = (0 + 0 + \frac{1}{2}\log(\frac{1}{2}) + \frac{1}{2}\log(\frac{1}{2}))$

$= 1$

$I(\text{ear Position}) = \frac{6}{8} \times 1.459 + \frac{2}{8} \times 1 = 1.344$

$\text{Information Gain} = 1.906 - 1.344 = \boxed{0.561}$

max gain with ear_shape → SPLIT 1

After split 1:

ear_shape = round (4 samples)

ear_shape = long (4 samples)

For ear_shape = round, (for 2nd split)

Nose :- (Showing calculations only for the best Split)

$H(\text{nose} = \text{small} \mid \text{ear\_shape} = \text{round}) = -(\frac{2}{2}\log(\frac{2}{2}) + 0 + 0 + 0)$

$= 0$

$H(\text{nose} = \text{big} \mid \text{ear\_shape} = \text{round}) = -(0 + 0 + 0 + \frac{2}{2}\log(\frac{2}{2}))$

$= 0$

$I(\text{nose} \mid \text{ear\_shape} = \text{round}) = \frac{1}{2}(0) + \frac{1}{2}(0) = 0$

Information gain = I(ear shape = round) - I(nose | ear shape = round)

= 1 - 0 = 1

ear_Position :- → Data = ear shape = long (4 samples)

$H(\text{ear\_Position} = \text{high} \mid \text{ear shape} = \text{long}) = -\left(\frac{3}{3}\log\left(\frac{3}{3}\right) + 0\right) = 0$

$H(\text{ear\_Position} = \text{low} \mid \text{ear Shape} = \text{long}) = -\left(0 + \frac{1}{1}\log\left(\frac{1}{1}\right)\right) = 0$

$I(\text{ear-Position} \mid \text{ear\_shape} = \text{long}) = \frac{3}{4}(0) + \frac{1}{4}(0)$

Information gain = I(ear shape = round)

I(ear-osition | ear_shape = long)

= 0.861

Purity after split 1

ear_shape = round → $\frac{2}{4} \times 100 = 50\%$

ear_shape = long → $\frac{3}{4} \times 100 = 75\%$

Purity after split 2

nose = Small | ear_shape = round = $\frac{2}{2} \times 100 = 100\%$

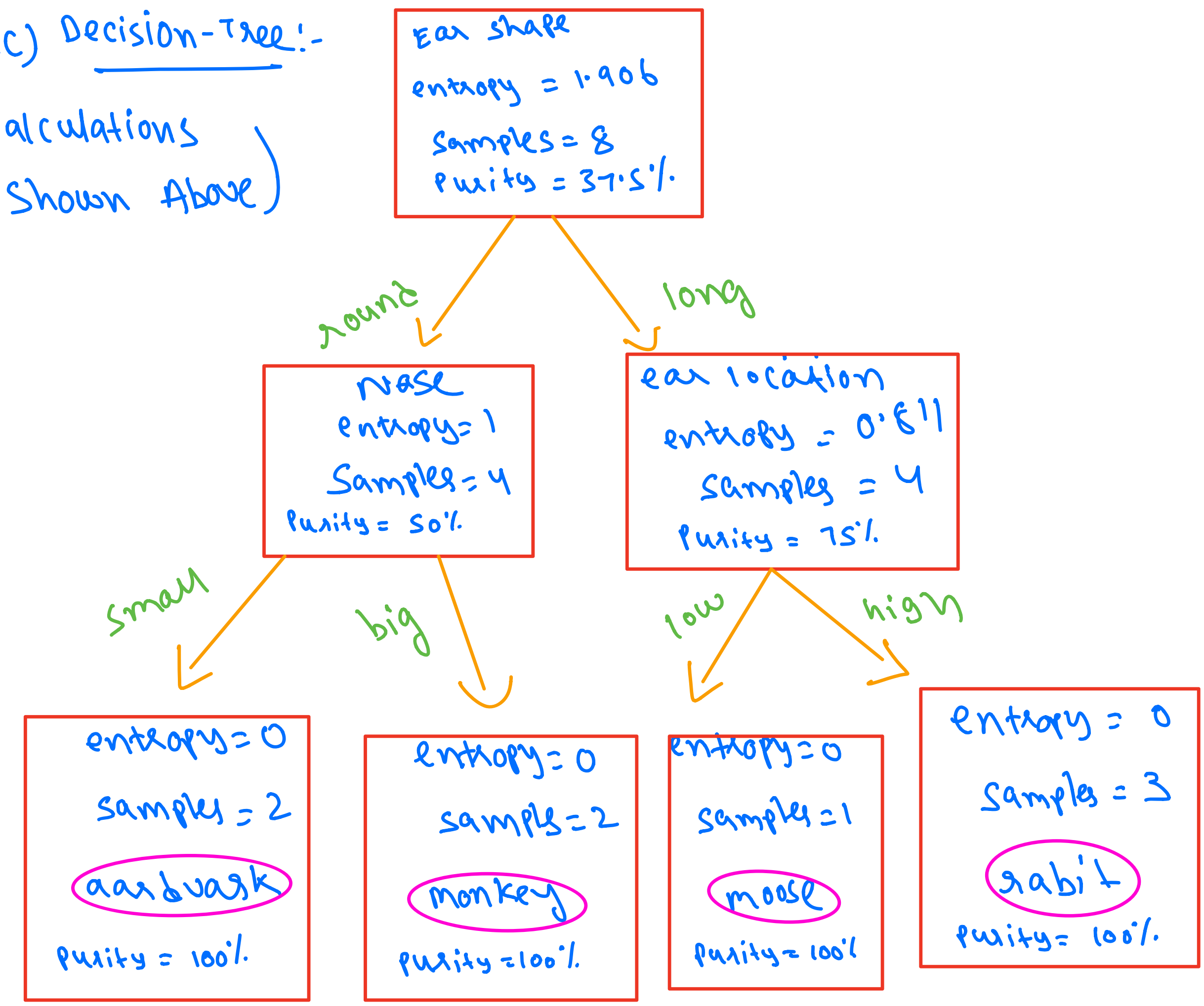
nose = big | ear_shape = round = $\frac{2}{2} \times 100 = 100\%$

ear_location = low | ear_Shape = long = $\frac{1}{1} \times 1 = 100\%$

ear_location = high | ear_shape = long = $\frac{3}{3} \times 1 = 100\%$

(c) Decision-Tree:-

(calculations Shown Above)



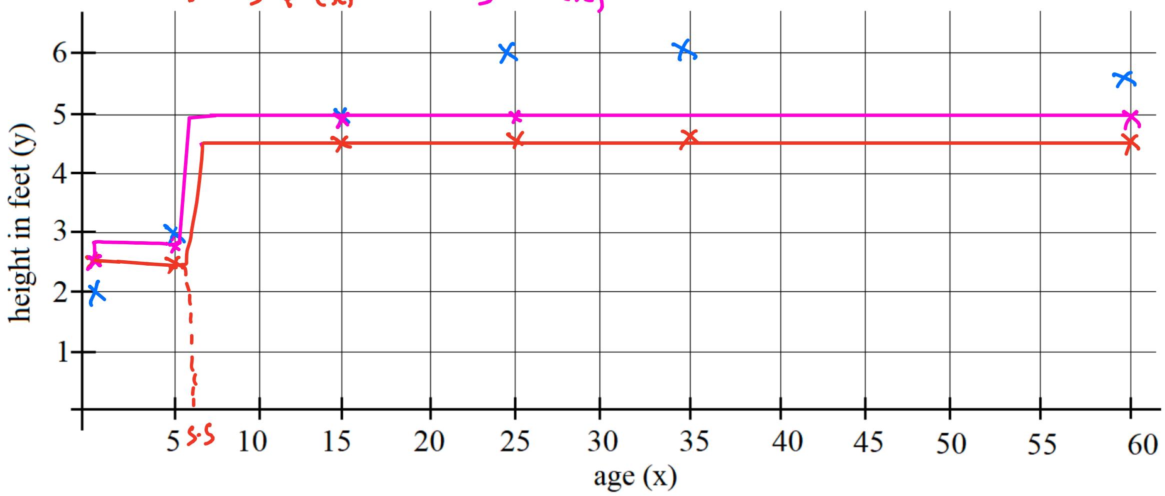
(d)

name	true label	inferred label (fill in)	nose	ear shape	ear location
Bambi	moose	Aardvark	small	round	high
Thumper	rabbit	moose	big	long	low
George	monkey	monkey	big	round	low

Test error $= \frac{2}{3}$

Q4) (a) · → Actual training data

· → $F^1(x)$ · → $F^2(x)$



(b) $f^{(0)}(x_i) = y_0 = \frac{1}{m} \sum_{i=1}^{m} y_i$

$$x = \begin{bmatrix} 25 \\ 60 \\ 15 \\ 1 \\ 5 \\ 35 \end{bmatrix} \quad f^{(0)}_x = \begin{bmatrix} 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \end{bmatrix} \quad y_{Actual} = \begin{bmatrix} 6 \\ 5.5 \\ 5 \\ 2 \\ 3 \\ 6 \end{bmatrix}$$

$$MSE = \frac{1}{6}\Big((6-4.58)^2 + (5.5-4.58)^2 + (5-4.58)^2 + (4.58-2)^2 + (4.58-3)^2 + (6-4.58)^2 \Big)$$

$$= 2.37$$ (nearest 0.01)

(c) Iteration 1 :-

(i) $f(x) = \begin{cases} c & a < x < b \\ 0 & \text{otherwise} \end{cases}$

$$\text{minimize} \sum_{i=1}^{m} \left[f(x_i) - z_i^{(0)} \right]^2 \quad — (1)$$

Optimization Problem :- we want to find a, b, c for values in x such that we minimize the function 1.

ii) Given , $a = 0$, $b = 5.5$

$$z^{(0)} = y_i - f_x^{(0)} = \begin{bmatrix} 1.417 \\ 0.917 \\ 0.417 \\ -2.583 \\ -1.583 \\ 1.417 \end{bmatrix}$$

$$f(x_i) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ c \\ c \\ 0 \end{bmatrix}$$

$$\text{minimize} \sum \left[(0 - 1.417)^2 + (0 - 0.917)^2 + (0 - 1.417)^2 + (c + 2.583)^2 + (c + 1.583)^2 + (0 - 1.417)^2 \right]$$

For minimum, take derivative and equate to 0.

$$\Rightarrow \quad 2(c + 2.583) + 2(c + 1.583) = 0$$

$$\boxed{c = -2.1}$$ nearest 0.1

iii) $F_x^{(1)} = f_x^{(1)} + f_x^{(0)}$

$$\Rightarrow \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2.1 \\ -2.1 \\ 0 \end{bmatrix} + \begin{bmatrix} 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \end{bmatrix} = \begin{bmatrix} 4.58 \\ 4.58 \\ 4.58 \\ 2.48 \\ 2.48 \\ 4.58 \end{bmatrix}$$

iv) $$\text{Loss} = \frac{1}{6}\left((6-4.58)^2 + (5.5-4.58)^2 + (5-4.58)^2 + (2-2.48)^2 + (3-2.48)^2 + (6-4.58)^2 \right)$$

$$= \boxed{0.93}$$ (nearest 0.01)

Residual $(2^1) \Rightarrow y_i - F^1(x) = \begin{bmatrix} 1.42 \\ 0.92 \\ 0.42 \\ -0.48 \\ 0.52 \\ 1.42 \end{bmatrix}$

Iteration 2 :-

i) $a = 1, \quad c = 0.4$

$$\min \sum \left(f^2(x) - \begin{bmatrix} 1.42 \\ 0.92 \\ 0.42 \\ -0.48 \\ 0.52 \\ 1.42 \end{bmatrix} \right)^2$$

Goal :-

$(\underbrace{f(x) - 2(i)}_{})$

we always wanted to reduce this, $|c - 2_i|$ should reduce.

This function would be minimized if we subtract c from all values except -0.58 (as $a = 1$).

It is guaranteed that the function would minimize and as values in $2^{(i)}$.

So, any $b > \max(x)$ would suffice

$\therefore b = 60.1$ (rounded off to nearest 0.1)

This is not generic, it depends on the value of c.

ii) $$F^2(x) = \begin{bmatrix} 0.4 \\ 0.4 \\ 0.4 \\ 0 \\ 0.4 \\ 0.4 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2.1 \\ -2.1 \\ 0 \end{bmatrix} + \begin{bmatrix} 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \\ 4.58 \end{bmatrix} = \begin{bmatrix} 4.98 \\ 4.98 \\ 4.98 \\ 2.48 \\ 2.88 \\ 4.98 \end{bmatrix}$$

ii) $\text{Loss} = \frac{1}{6}\Big[(6-4.98)^2 + (5.5-4.98)^2 + (5-4.98)^2 + (2-2.48)^2 + (3-2.88)^2 + (6-4.98)^2\Big]$

$= \boxed{0.43}$ (nearest 0.01)

Residual $2^{(2)} \Rightarrow y - F_2^{(x)} = \begin{bmatrix} 1.02 \\ 0.52 \\ 0.02 \\ -0.48 \\ 0.12 \\ 1.02 \end{bmatrix}$

e)

person	age	true height	fitted height
Gary	12	5	4.98
Harris	57	5	4.98
Ivy	3	2	2.88

Test mean squared error

$\Rightarrow \frac{1}{3}\Big[(5-4.98)^2 + (5-4.98)^2 + (2-2.88)^2\Big]$

$= \boxed{0.26}$

5)

(a)

points in cluster 1	points in cluster 2	points in cluster 3
$x_2 = 1$	$x_9 = 2$	$x_1 = 6$
$x_7 = 1$		$x_3 = 12$
$x_8 = 1$		$x_4 = 14$
		$x_5 = 3$
		$x_6 = 5$

(b) $C_1 = \frac{1+1+1}{3} = 1$

$C_2 = \frac{2}{1} = 2$

$C_3 = \frac{6+12+14+3+5}{5} = 8$

$\boxed{C_1 = 1, \quad C_2 = 2, \quad C_3 = 8}$

(c) $A = N(30, 25)$ $\qquad N(\mu, \sigma^2)$

(notation)

$B = N(50, 5)$

$C = N(60, 100)$

$A = 3(B + C)$ $\Rightarrow$

$B = 4C$

$$\boxed{P(A) = \frac{15}{20} \quad P(B) = \frac{4}{20} \quad P(C) = \frac{1}{20}}$$

Combined Normal:-

$$\mu' = P_A \mu_A + P_B \mu_B + P_C \mu_C$$

$$= \frac{15}{20} \times 30 + \frac{4}{20} \times 50 + \frac{1}{20} \times 60$$

$$= 35.5$$

$$(\sigma')^2 = (P_A^2 \sigma_A^2) + (P_B^2 \sigma_B^2) + (P_C^2 \sigma_C^2)$$

$$= \left(\frac{15}{20}\right)^2 \times 25 + \left(\frac{4}{20}\right)^2 \times 5 + \left(\frac{1}{20}\right)^2 \times 100$$

$$= 14.51$$

$$\boxed{(\mu', \sigma'^2) = (35.5, 14.51)}$$

$$\boxed{f(x) = \frac{1}{\sqrt{2 * \pi * 14.51}} e^{-\frac{(x-35.5)^2}{14.51}}}$$

(d) $f(x = 40cm)$,

Applying Baye's Rule;

$P(\text{type} = \text{Ajax} \mid \text{size} = 40) =$

$$\Rightarrow \frac{0.75 \times (N(30,25)\ @\ 40cm)}{0.75 \times N(30,25)\ @\ 40cm) + 0.2 \times (N(50,5)\ @\ 40\ cm) + 0.05 \times N(60,100)\ @\ 40cm)}$$

used python

$\Rightarrow$ 0.97

$$\boxed{\therefore P(\text{type} = \text{Ajanc} \mid \text{size} = 40) = 0.97}$$

e) P(you are poisoned | size = 50)

A C

$$\Rightarrow \frac{(0.75 \times N(30,25)\ @\ 50cm) + (0.05 \times N(60,100)\ @\ 50cm)}{(0.75 \times N(30,25)\ @\ 50cm) + (0.2 \times N(50,5)\ @\ 50cm)) + (0.05 \times N(60,100)\ @\ 50cm)}$$

$\Rightarrow$ 0.03

$$\boxed{\therefore P(\text{I have been poisoned} \mid \text{size} = 50cm) = 0.03}$$

In [233]:
```python
import math
import numpy as np
def getNormalPDF(x, mean, sd):
    var = float(sd)**2
    denom = (2*math.pi*var)**.5
    num = math.exp(-(float(x)-float(mean))**2/(2*var))
    return num/denom
```

In [234]:

(sd)

```python
a = getNormalPDF(40,30,np.sqrt(25))
b = getNormalPDF(40,50,np.sqrt(5))
c = getNormalPDF(40,60,np.sqrt(100))
p = 0.75*a/((0.75*a) + (0.2*b) + (0.05*c))
round(p,2)
```

Out[234]: 0.97

In [235]:

(se)

```python
a = getNormalPDF(50,30,np.sqrt(25))
b = getNormalPDF(50,50,np.sqrt(5))
c = getNormalPDF(50,60,np.sqrt(100))
p = (0.75*a)  + (0.05*c)/((0.75*a) + (0.2*b) + (0.05*c))
round(p,2)
```

Out[235]: 0.03

b)

A	B	C
T	T	T
T	T	F
T	F	T
F	T	T

$$P = (0.2)^3 + 3 \times (0.2)^2 (0.8)$$

$$= 0.08 + 0.96 = 0.104$$

$$\boxed{P(\text{answer will be true}) = 0.104}$$

6) (b) $S_{MLE} = P(\text{answer is } \hat{s} \mid A, B, C)$

$\hat{s} \in \{T, F\}$

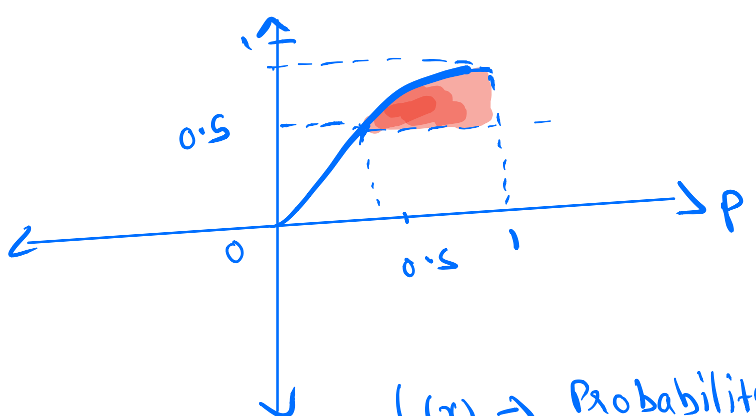
correct answer → each p

$$P(\text{majority voting is correct}) = \underbrace{p^3}_{\substack{\text{All of them} \\ \text{are correct}}} + \underbrace{3p^2(1-p)}_{\substack{\text{at least 2} \\ \text{of them} \\ \text{are correct}}}$$

$p^3 + 3p^2(1-p) > 0.5$

is strictly increasing $\left(\frac{1}{2}, 1\right]$

$\therefore$ if $p > \frac{1}{2}$, $\hat{S}_{MLE}$ is the majority vote



$f(x) \Rightarrow$ Probability mass function

c) $E[x] = x f(x)$

$E[x^2] = x^2 f(x)$, $\operatorname{var}(x) = E[x^2] - E(x)^2$

d) $\hat{S}_{MAP} = 1 \rightarrow T \rightarrow 1-(1-p)^3$

$\hat{S}_{MAP} = 0 \rightarrow (1-p)^3$

This is nothing but pmf(x)

$$E[S_{MAP}] = (1-(1-p)^3) \times 1 + [(1-p)^3 \times 0]$$

$$\boxed{\text{mean} = 1-(1-p)^3}$$

$$var(x) = E(x^2) - E(x)^2$$

$$E(x^2) = 1^2 (1-(1-p)^3 + 0^2 (1-p)^3$$

$$= (1-(1-p)^3)$$

$$var(x) = (1-(1-p)^3) - (1-(1-p)^3)^2$$

$$= (1-(1-p)^3)[1-(1-(1-p)^3]$$

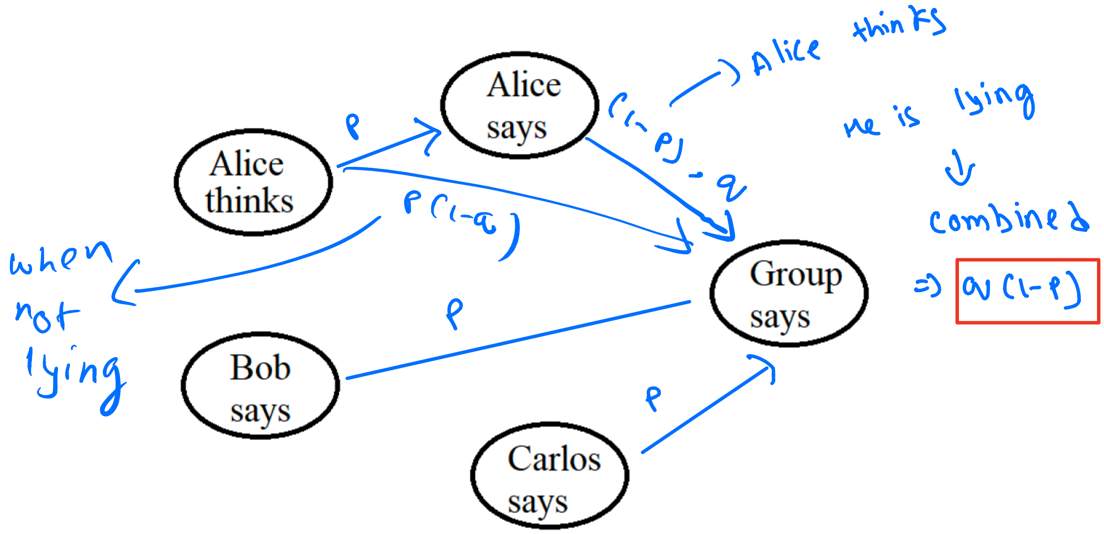
$$= (1-(1-p)^3)[(1-p)^3]$$

$$var(x) = (1-p)^3 (1-(1-p)^3)$$

$$\boxed{\therefore \text{var} = (1-p)^3 (1-(1-p)^3)}$$

e)

i)



ii) $P(\text{correct answer}) = P(\text{correct answer} \mid A \text{ is Sabotaging})$

$+ P(\text{correct Answer} \mid A \text{ is not Sabotaging})$

These are the only cases, we get a right answer

A	B	C
1	1	1
1	1	0
0	1	1
1	0	1

$$\Rightarrow (0.3a \times 0.7 \times 0.7 + 0.3a \times 0.7 \times 0.3 + 0.7a \times 0.7 \times 0.7 + 0.3a \times 0.3 \times 0.7)$$

$$+ (0.7 \times (1-a) \times 0.7 \times 0.7 + 0.7 \times (1-a) \times 0.7 \times 0.3 + 0.3 \times (1-a) \times 0.7 \times 0.7 + 0.7 \times (1-a) \times 0.3 \times 0.7)$$

$$\Rightarrow \cancel{0.7} = a(0.3 \times \cancel{0.7} \times 0.7 + 0.3 \times \cancel{0.7} + 0.3 + \cancel{0.7} \times 0.7 \times 0.7 + 0.3 \times 0.3 \times \cancel{0.7}) +$$

$$(1-a)(0.7 \times \cancel{0.7} \times 0.7 + \cancel{0.7} \times 0.7 \times 0.3 + 0.3 \times \cancel{0.7} \times 0.7 + \cancel{0.7} \times 0.3 \times 0.7)$$

$$\Rightarrow 1 = w\,(0.21 + 0.09 + 0.49 + 0.09) + (1-w)\,(0.49 + 0.21 + 0.21 + 0.21)$$

$$1 = w\,(0.88) + (1-w)\,(1.12)$$

$$1 = 0.88w + 1.12 - 1.12w$$

$$0.24w = 0.12$$

$$\boxed{w = 0.5}$$