Question 1:

1a.

$$f(W) = P(\frac{Y}{X} \leq W) \quad A = X \geq 0 \; and \; A = X \; < 0$$

$$= P\,(\frac{Y}{X} \leq W, A \;\cup \bar{A})$$

$$= P\,\left(\frac{Y}{X} \leq W, X \;\geq 0\right) + \;P\,\left(\frac{Y}{X} \leq W, X < \;0\right)$$

=

$$= P\,(Y \leq WX, X \;\geq 0) + \;P\,(Y \;\leq WX, X < \;0)$$

$$f(W) = \int_0^\infty \int_{-\infty}^{XW} f_{YX}(x,y)dxdy \;+ \int_{-\infty}^0 \int_{XW}^\infty f_{YX}(x,y)dxdy$$

As, $X \;\geq 0\;, Y \;\geq 0, we\ will\ use\ only\ the\ first\ part\ of\ the\ above\ equation\ for\ our\ PDF.$
By expanding the first part we get the below,

$$f(W) = \int_0^\infty x\; f_{XY}(XW, X)dx$$

$$= \int_0^\infty x\;\; f_{XY}(XW, X)dx$$

$$= \int_0^\infty x\;\lambda\mu e^{-\lambda x - \mu y} dx$$

$$= \int_0^\infty x\;\lambda e^{-\lambda x(1-e^{-\mu x})}\; dx$$

$$= \int_0^\infty x\;\lambda e^{-\lambda x}\; dx - \int_0^\infty \mu e^{-(\lambda+\mu)x}\; dx$$

$$= 1 - \frac{\lambda}{\lambda + \;\mu}\int_0^\infty x(\lambda+\mu)e^{-(\lambda+\mu)x}\; dx$$

Replacing, $x(\lambda + \mu) = u$,

$$= 1 - \frac{\lambda}{\lambda + \;\mu}\;\int_0^\infty ue^{-u}\; du$$

As, $\int_0^\infty ue^{-u}\; du \;\; = 1,$

$$= 1 - \frac{\lambda}{\lambda + \mu}$$

$$= \frac{\lambda + \mu - \lambda}{\lambda + \mu}$$

$$= \frac{\mu}{\lambda + \mu}$$

1b.

Assumption:

X and Y are exponentially distributed random variables with parameters $\lambda\ and\ \mu$

$$f_{XY}(x, y) = \begin{cases} \lambda\mu e^{-(\lambda x + \mu y)}, x \geq 0, y \geq 0 \\ 0, \quad otherwise \end{cases}$$

$$f_{XY}(x, y) = \lambda e^{-\lambda x}. \mu e^{-\mu y} = \lambda\mu e^{-\lambda x - \mu y}$$

$$P(X < Y) = \iint_{x<y}^{\infty} \lambda\mu e^{-\lambda x - \mu y} dxdy$$

$$= \int_{x=0}^{\infty} dx \int_{y=x}^{\infty} \lambda\mu e^{-\lambda x - \mu y} dy$$

$$= \int_{x=0}^{\infty} \lambda e^{-\lambda x - \mu y} dx$$

$$= \frac{\lambda}{\lambda + \mu}$$

Question 2:

2a.

Majority voting of the ensemble method is obtained by combining the three independent binary classifiers in such a way as to result a superior model which gives better accuracy over classification but this increases the overall error rate compared to the individual classifiers.

(Ci,i=1,2,3)

- Calculate the error rate of each model

Error rate = 1 – Accuracy = No. of Incorrect classifications / Total data points

- Assigns higher weights to model that has more errors
- Take mode on the error rates

As majority voting classifier is based on the assumption that all the three binary classifier has a same error rate, this gives better classification.

Using majority voting ensemble for error rate, we will get 0.3 as our result as both e2 and e3 has values of 0.3

2b.

The assumption of independence is relaxed on the errors, means that, we have the target value available our analysis.
With more training data points, contingency table, assumption of independence on the errors can be studied

When the assumption of independence is relaxed on error rate, we get better error rate, as now we consider the classifier has joint probabilities.

Question 3:

ID	X1	X2	X3	X4	Y
Training					
1	S	42.5	N	F	N
2	S	39.2	H	T	N
3	-	33.6	H	F	Y
4	R	-	H	F	Y
5	R	22.8	N	F	Y
6	R	15.4	N	T	N
Testing					
7	O	25.0	N	T	?
8	S	36.4	-	F	?

Preprocessing:

- Impute random value for X1 from (S,R)
    - Impute S for record 3
    - Change O to R for record 7
- Impute mean value for X2
- Impute random value from (N,H) for record 8

Assumptions:

- X1, X3 and X4 has only two classes (eg. S or R for X1)
- For a given particular value of Y, the distribution of one feature is independent of other three

ID	X1	X2	X3	X4	Y

Training					
1	S	42.5	N	F	N
2	S	39.2	H	T	N
3	**S**	33.6	H	F	Y
4	R	**25.6**	H	F	Y
5	R	22.8	N	F	Y
6	R	15.4	N	T	N
Testing					
7	**R**	25.0	N	T	?
8	S	36.4	**N**	F	?

Now our data is clean and is ready for model building.
Naïve Bayes is a supervised classification model which is build based on the Bayes theorem. Mixed Naïve Bayes is used when the dataset has both categorical and continuous features. This situation can be handled in two ways:

Method 1:

a. Convert continuous feature into categorical by grouping values into bins (either ordinal or nominal) e.g. age(0-5,5-10 etc.)
b. Create one final multinomial model to predict the output

Method 2:

c. Create two separate models (categorical or multinomial classification model and continuous gaussian model for respective features)
d. ensemble the probabilities of the above created 2 models and build a gaussian model on those probabilities
e. predict the output from the ensemble model

As, the question mentioned, we need to use both gaussian and multinomial models, I took Method 2 for my explanation.

Pre-Calculations:

From the 6 training datapoints, I estimated the following:

$$p(y) = \begin{cases} \frac{3}{6} & if\ y = Y \\ \frac{3}{6} & if\ y = N \end{cases}$$

Out of 6 training datapoints, 3 target y value is ‘Y’ and 3 target y value is ‘N’.

$$p(X1|y = Y) = \begin{cases} \frac{1}{3} & if\ X1 = S \\ \frac{2}{3} & if\ X1 = R \end{cases}$$

For column X1, with 3 target value Y, there are 1 S value, and 2 R value.
Similarly calculated for X3 and X4 for ‘Y’ and for X1, X3, and X4 for target value ‘N’

$$p(X3|y = Y) = \begin{cases} \frac{1}{3} & if\ X3 = N \\ \frac{2}{3} & if\ X3 = H \end{cases}$$

$$p(X4|y=Y) = \begin{cases} \frac{0}{2} \ if \ X4 = T \\ \frac{3}{4} \ if \ X4 = F \end{cases}$$

$$p(X1|y=N) = \begin{cases} \frac{2}{3} \ if \ X1 = S \\ \frac{1}{3} \ if \ X1 = R \end{cases}$$

$$p(X3|y=N) = \begin{cases} \frac{2}{3} \ if \ X3 = N \\ \frac{1}{3} \ if \ X3 = H \end{cases}$$

$$p(X4|y=N) = \begin{cases} \frac{2}{2} \ if \ X4 = T \\ \frac{1}{4} \ if \ X4 = F \end{cases}$$

Implement Bayes Theorem:

$$\boldsymbol{p(y|x) = \frac{p(x|y)\,.\,p(y)}{p(x)}}$$

$$p(y=N|X1=S,X3=N,X4=F) = \frac{\big(p(X1=S,X3=N,X4=F|y=N)\,.\,p(y=N)\big)}{p(X1=S,X3=N,X4=F)}$$

$$= \frac{\big(p(X1=S,X3=N,X4=F|y=N)\,.\,p(y=N)\big)}{\sum_{i=Y,N} p(X1=S,X3=N,X4=F\ |y=i)\,.\,p(y=i)}$$

$$= \frac{\big(p(X1=S|y=N)\,.\,p(\ X3=N|y=N)\,.\,p(X4=F|y=N)\,.\,p(y=N)\big)}{\sum_{i=Y,N} p(X1=S\ |y=i)\,.\,p(\ X3=N|y=i)\,.\,p(X4=F|y=i)\,.\,p(y=i)}$$

Now, I calculate for each individual component in the above formula

$p(X1=S\|y=Y)$	1/3
$p(\ X3=N\|y=Y)$	1/3
$p(X4=F\|y=Y)$	3/4
$p(X1=S\|y=N)$	2/3
$p(\ X3=N\|y=N)$	2/3
$p(X4=F\|y=N)$	1/4
$p(y=Y)$	3/6
$p(y=N)$	3/6

Substituting the derived values into the Bayes formula, I got the below table,

Probability Formula	Value
$p(y=N\|X1=S,X3=N,X4=F)$	0.5717
$p(y=N\|X1=S,X3=H,X4=T)$	1
$p(y=Y\|X1=S,X3=H,X4=F)$	0.75
$p(y=Y\|X1=R,X3=H,X4=F)$	0.3472

$p(y = Y \mid X1 = R, X3 = N, X4 = F)$	0.75
$p(y = N \mid X1 = R, X3 = N, X4 = T)$	1

The above table gives the output probabilities for multinomial (categorical) model.

Next, I calculate the gaussian model for the continuous feature i.e. X2 with Y as my target value for the given 6 training data points.

Pre-Calculations:

I calculated the distributions using continuous data,

$$p(y) = \begin{cases} \frac{3}{6} & if\ y = Y \\ \frac{3}{6} & if\ y = N \end{cases}$$

$p(X2 \mid y = Y) = \frac{1}{2\pi\sigma} \exp\left(-\frac{(X2 - \mu)^2}{2\sigma}\right)$,
where μ *is the mean and* σ *is the standard deviation*
Calculated mean and standard deviation for X2 with target value Y and N respectively,

$$\mu_Y = 27.34, \sigma_Y = 4.58, \mu_N = 32.37, \sigma_N = 12.07$$

$p(X2 = 42.5 \mid y = N)$	0
$p(X2 = 39.2 \mid y = N)$	0.00191
$p(X2 = 33.6 \mid y = Y)$	0.00045
$p(X1 = 25.6 \mid y = Y)$	0.0253
$p(X3 = 22.8 \mid y = Y)$	0.00364
$p(X2 = 15.4 \mid y = N)$	0

Applying Bayes theorem for the probabilities I got the below table,

Probability Formula	Value
$p(X2 = 42.5 \mid y = N)$	0
$p(X2 = 39.2 \mid y = N)$	1
$p(X2 = 33.6 \mid y = Y)$	0.0352
$p(X1 = 25.6 \mid y = Y)$	0.9280
$p(X3 = 22.8 \mid y = Y)$	0.9262
$p(X2 = 15.4 \mid y = N)$	0

Next I build an ensemble gaussian model for the two derived set of probabilities against the target value

Feature_1_prob_cat (Z1)	Feature_2_prob_cnt (Z2)	Target value
0.5717	0	N
1	1	N
0.75	0.0352	Y
0.3472	0.9280	Y

0.75	0.9262	Y
1	0	N

Calculate mean and standard deviation for the above features,

Feature_1_prob_cat
(Z1):

$$\mu_Y = 0.6157, \sigma_Y = 0.1898, \mu_N = 0.8572, \sigma_N = 0.2019$$

Feature_2_prob_cnt
(Z2):

$$\mu_Y = 0.6298, \sigma_Y = 0.4204, \mu_N = 0.333, \sigma_N = 0.4714$$

$p(Z1 = 0.5717 \mid y = N)$	0.6445
$p(Z1 = 1 \mid y = N)$	0.7498
$p(Z1 = 0.75 \mid y = Y)$	0.7995
$p(Z1 = 0.3472 \mid y = Y)$	0.6935
$p(Z1 = 0.75 \mid y = Y)$	0.7995
$p(Z1 = 1 \mid y = N)$	0.7498
$p(Z1 = 0.5717 \mid y = Y)$	0.8343
$p(Z1 = 1 \mid y = Y)$	0.5684
$p(Z1 = 0.75 \mid y = N)$	0.7665
$p(Z1 = 0.3472 \mid y = N)$	0.4141
$p(Z1 = 0.75 \mid y = N)$	0.7665
$p(Z1 = 1 \mid y = Y)$	0.5684
$p(Z2 = 0 \mid y = N)$	0.3002
$p(Z2 = 1 \mid y = N)$	0.2108
$p(Z2 = 0.0352 \mid y = Y)$	0.2487
$p(Z2 = 0.9280 \mid y = Y)$	0.3407
$p(Z2 = 0.9262 \mid y = Y)$	0.3411
$p(Z2 = 0 \mid y = N)$	0.3002
$p(Z2 = 0 \mid y = Y)$	0.3328
$p(Z2 = 1 \mid y = Y)$	0.3217
$p(Z2 = 0.0352 \mid y = N)$	0.3073
$p(Z2 = 0.9280 \mid y = N)$	0.2321
$p(Z2 = 0.9262 \mid y = N)$	0.2326
$p(Z2 = 0 \mid y = Y)$	0.3328

Substitute the above values into Bayes theorem to derive the final probability for ensembled gaussian model

Feature_1_prob_cat (Z1)	Feature_2_prob_cnt (Z2)	Target value	Probability
0.5717	0	N	0.4106
1	1	N	0.4636
0.75	0.0352	Y	0.4578
0.3472	0.9280	Y	0.7108
0.75	0.9262	Y	0.6047
1	0	N	0.5433

Now, apply the formula to find the target value on test data

Testing					
7	**R**	25.0	N	T	?
8	S	36.4	**N**	F	?

Based on the pre-calculations for categorical model, we substitute for Bayes theorem and get the below,

Probability Formula	Value
$p(y = Y \| X1 = R, X3 = N, X4 = T)$	0
$p(y = Y \| X1 = S, X3 = N, X4 = T)$	0.4285

Now, calculate probability for continuous feature

Probability Formula	Value
$p(y = Y \| X2 = 25.0)$	0.0191
$p(y = Y \| X2 = 36.4)$	0
$p(y = N \| X2 = 25.0)$	0.00139
$p(y = N \| X2 = 36.4)$	0.00673

Apply Bayes theorem,

Probability Formula	Value
$p(y = Y \| X2 = 25.0)$	0.932
$p(y = Y \| X2 = 36.4)$	0
$p(y = N \| X2 = 25.0)$	0
$p(y = N \| X2 = 36.4)$	0.4285

Bayes theorem for ensembled model on test data,

Testing					**PREDICTED PROBABILITY**	**PREDICTED _TARGET**
7	**R**	25.0	N	T	0.4520	Y
8	S	36.4	**N**	F	0.6045	Y

Final prediction gives Y value for both test data point.