1.1.

$P(B / A)P(A) = P(B \cap A)$
$P(B / \neg A)P(\neg A) = P(B \cap \neg A)$
$P(B \cap S) = P(B \cap A) + P(B \cap \neg A) = P(B)$

Now from above we can can write the given equation as

$$P(A / B) = \frac{P(B / A)P(A)}{P(B)}$$

$$P(A / B)P(B) = P(B / A)P(A)$$

This is Bayes rule. Hence we can say that

$$P(A / B) = \frac{P(B / A)P(A)}{P(B \cap A) + P(B \cap \neg A)}$$

1.2.a

Two variables X and Y are independent if they satisfy the condition $P(X \cap Y) = P(X)P(Y)$

$P(X = 0) = 0.575$
$P(X = 1) = 0.425$
$P(Y = 0) = 0.35$
$P(Y = 1) = 0.65$

$P(X=0 \cap Y=0) = 0.2 \cong P(X = 0)P(Y = 0)$
$P(X=0 \cap Y=1) = 0.375 \cong P(X = 0)P(Y = 1)$
$P(X=1 \cap Y=0) = 0.15 \cong P(X = 1)P(Y = 0)$
$P(X=1 \cap Y=1) = 0.275 \cong P(X = 1)P(Y = 1)$

We can see that $P(X \cap Y) \cong P(X)P(Y)$ Hence we can say that X and Y are independent.

1.2.b

Conditional Independence can be stated by the condition
$P(A \cap B / C) = P(A / C) * P(B / C)$

In this question we were asked if
$P(X \cap Y / Z) = P(X / Z) * P(Y / Z)$

$P(X=0 \cap Y=0/Z=0) = P(X=0,Y=0,Z=0) / P(Z=0)$
$= 0.1 / 0.45$
$= 0.222$

$$P(X=0 / Z=0) * P(Y=0 / Z=0) = \frac{P(X=0,Z=0) * P(Y=0,Z=0)}{P(Z=0) * P(Z=0)}$$
$$= \frac{0.3 * 0.15}{0.45 * 0.45}$$
$$= 0.222$$

$P(X=1 \cap Y=1/Z=1) = P(X=1,Y=1,Z=1) / P(Z=1)$
$= 0175/0.65$
$= 0.269$

$$P(X=1 / Z=1) * P(Y=1/ Z=1) = \frac{P(X=1,Z=1) * P(Y=1,Z=1)}{P(Z=1) * P(Z=1)}$$
$$= \frac{0.275 * 0.35}{0.65 * 0.65}$$
$$= 0.2278$$

$P(X=1 \cap Y=1/Z=1) \neq P(X=1 / Z=1) * P(Y=1/ Z=1)$

Hence we can say that X is not conditionally independent of Y given Z.

1.2.c

P(X ≠ Y / Z=0) = (P(X=0,Y=1 / Z=0) + P(X=1,Y=0 / Z=0)) / P(Z=0)
= (0.2 + 0.05) / 0.45
= 0.555

2.1

Let us assume there is X = $\{x_1 \ldots\ldots x_n\}$ where $x_i \in \{0,1\}$

We have L(θ) = $\prod_{i=1}^{n} \theta^{x_i} * (1-\theta)^{1-x_i}$

l(θ) = log(L(θ))
= log($\prod_{i=1}^{n} \theta^{x_i} * (1-\theta)^{1-x_i}$)
= $\sum_{i=1}^{n} x_i \log\theta + (1 - x_i)\log(1 - \theta)$

The Log likelihood function does not depend on the order of the random variable. It only depends on the no of ones in the random variable.

2.2

Now let us try to find out the derivative of the log likelihood function with respect to θ. It can be written as.

l'(θ) = $\frac{\sum_{i=1}^{n}(x_i-\theta)}{\theta(\theta-1)}$

Equation l'(θ) = 0, we get

$\boldsymbol{\theta} = \frac{\sum_{i=1}^{n} \mathbf{x_i}}{\mathbf{n}}$

if we have k = $\sum_{i=1}^{n} x_i$ = no of 1's in data

For the ten samples the maximum likelihood estimate would be for $\theta = k/n$ = 6/10 = 0.6 , substituting in the likelihood equation we get,

L(θ) = $\theta^k(1-\theta)^{n-k} = 0.6^6(1-0.6)^4$ = **0.072**

2.3

The likelihood is given by

L(θ) = $\prod_{i=1}^{m} \frac{n!}{k_i!(n-k_i)!} \theta^{k_i}(1-\theta)^{n-k_i}$

When we do log likelihood we will get

l(θ) = $\sum_{i=1}^{m} log\left(\frac{n!}{k_i!(n-k_i)!} \theta^{k_i}(1-\theta)^{n-k_i}\right)$

l(θ) = $\sum_{i=1}^{m} log\left(\frac{n!}{k_i!(n-k_i)!}\right) + k_i log\,\theta + (1-k_i)log(1-\theta)$

2.4

Computing the derivative of log likelihood from the above question. We get

$l'(\theta) = \sum_{i=1}^{m} \frac{n\theta - k_i}{\theta(\theta-1)}$

Equating it to zero we have again

$$\boldsymbol{\theta = \frac{\sum_{i=1}^{m} k_i}{mn}}$$

Maximum likelihood estimate of $\theta = \frac{\sum_{i=1}^{m} k_i}{mn}$, using it in the likelihood equation we get

$\theta = 6/10 = 0.6$

$P(Y_1 = 3) = \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k} = \frac{5!}{3!(2)!}0.6^3(0.4)^2 = 0.376$

$P(Y_2 = 3) = \frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k} = \frac{5!}{3!(2)!}0.6^3(0.4)^2 = 0.376$

Maximum likelihood estimate = 0.376 * 0.376 = **0.141**

2.5

Each Y_i represents part1's X = $\{x_1 \ldots\ldots x_n\}$ and all Y_i combined together represents part3. We can see similar formulas for part1 and part3, the difference is that in part 3 we are considering 2D X and in part 1 we have a 1D X. The optimal θ we got for both part 2 and part 4 is different, this is because of the dimensionality difference, in part 1 its one dimensional and part3 its two dimensional. But if we look at θ we get optimal θ as the ratio of (no of 1's/ all points in data) for both the cases. For part 1, $\theta = k/n$ which is ratio of (no of 1's/ all points in data) and for part 3, $\theta = \frac{\sum_{i=1}^{m} k_i}{mn}$ which is also ratio of (no of 1's/ all points in data). Maximum likelihood estimate of part4 is greater than the part2, even tough optimal θ for them is almost the same, since part4 is two dimensional and part2 is one dimensional and probability formula is different for both. In part3 and part4 we are also worried about different combinations possible that gives the $\frac{n!}{k!(n-k)!}$ term. In part 1 and part 2 we are not worried about all the combinations possible we have one fixed combination.

3.

The code for the question is uploaded in canvas

```
Training Accuracy Obtained for NB is  0.9756959787892179
Test Accuracy Obtained for NB is  0.971191135734072
Training time required for NB is 0.07139706611633301
```

I have used Multinomial Naive Bayes and gave alpha =1 for smoothing. With the help of smoothing we can avoid the 0 probability conditions.