Q3

(a) A: adversal pertubations

B: Backdoor attacks

m: Misclassification alarm

Posterior Probability $P(A|M) = \frac{P(M|A) \cdot P(A)}{P(M)}$

$P(M|A)$ → likelihood of misclassification given adversal pertubation occured

$P(A)$ → Prior probability of adversal pertubation

$P(M)$ → Total probability of misclassification

here $P(M) = P(M|A) \cdot P(A) + P(M|B) \cdot P(B)$

(b) Prior Probabilities are:

- $P(A)$: initial/Prior belief of adversal pertubations
- $P(B)$: initial/Prior belief of backdoor attacks.

Likelihood Probabilities are:

- $P(M|A)$: Probability of observing misclassification given that adversal pertubation has occcured.
- $P(M|B)$: Probability of observing misclassificath given that backdoor attack has occcured

Posterior Probability:

$P(A|M)$: Probability/belief of likelihood of adversal pertubation given that a misclassification has occured

(c)

- it is given that there is an increase in backdoor attacks, which results in updation of prior $P(B)$ to increas.
- since $P(M)$ is directly positively related to $P(B)$.

using part (a) : $P(M) = P(M|A) \cdot P(A) + P(M|B) \cdot P(B)$

increase in probability of backdoor attack will result in an increase in probability of a misclassification alarm

- $P(A|M) = \dfrac{P(M|A) \cdot P(A)}{P(M|A) \cdot P(A) + P(C|B) \cdot P(B)}$

An increase in $P(B)$ will result in reduction of our belief regarding the likelihood of adversal perturbation causing the misclassification due to inverse positive relationship.