

**Question no 1:** Suppose you have used your favourite concept learning algorithm to learn a hypothesis  $h_1$  from some training data. You are interested in knowing the accuracy that the hypothesis can be expected to achieve on the underlying population. To assess this accuracy you apply the hypothesis to a test data set consisting of 45 instances that you had held back from the training data set. The error rate observed on the training data is 6.67%. Calculate the 95% confidence interval for the true error?

- $n$  (number of instances) = 45
- $e$  (error rate) = 6.67% = 0.0667

The 95% confidence interval for the true error 'e' is:

$$e = e \pm 1.96 * \text{sqrt}((e * (1 - e)) / n)$$

$$\begin{aligned}\text{Standard Error} &= \text{sqrt}((0.0667 * (1 - 0.0667)) / 45) \\ &= \text{sqrt}((0.0667 * 0.9333) / 45) \\ &= \text{sqrt}(0.06220011 / 45) \\ &= \text{sqrt}(0.0013822246) \\ &= 0.037178\end{aligned}$$

$$\begin{aligned}\text{Margin of Error} &= 1.96 * 0.037178 \\ &= 0.07286888\end{aligned}$$

$$\begin{aligned}\text{Lower bound} &= 0.0667 - 0.07286888 \\ &= -0.00616888\end{aligned}$$

$$\begin{aligned}\text{Upper bound} &= 0.0667 + 0.07286888 \\ &= 0.13956888\end{aligned}$$

Since an error rate can't be negative, we can say the lower bound is 0.

So, 95% confidence interval for the true error is approximately **0 to 0.1396** means **0% to 13.96%**.