Question 1 (10 points):

Consider a learned hypothesis, h, for some Boolean concept. When h is tested on a set of 100 examples, it classifies 83 correctly. What is the standard deviation and the 95% confidence interval for the true error rate for ErrorD(h)?

① standard deviation $S = \sqrt{\sigma^2} = \sqrt{np(1-p)}$

$$= \sqrt{100 \times \left(\frac{100-83}{100}\right) \times \left(\frac{83}{100}\right)}$$

$$= \sqrt{14.11} \approx 3.76$$

standard deviation estimate for $Error_D(h) = \frac{3.76}{100} = 0.0376$

$= 3.76\%$

② 95% confidence interval for ~~Error(h)~~ $Error_D(h)$

$$= error_D(h) \pm z_N \sqrt{\frac{error_D(h)(1-error_D(h))}{n}}$$

(z-score for 95% confidence level is 1.96)

$$= 0.17 \pm 1.96 \times \sqrt{\frac{0.17 \times (1-0.17)}{100}} = 0.0964 \sim 0.2436$$

Question 2 (10 points):

Consider again the example application of Bayes rule in Section 6.2.1 of Tom Mitchell's textbook (or slide page 6 of Lecture 7). Suppose the doctor decides to order a second laboratory test for the same patient, and suppose the second test returns a positive result as well. What are the posterior probabilities of cancer and ¬cancer following these two tests? Assume that the two tests are independent.

we can know if both two test are positive, the posterior probability of cancer and ¬cancer are $P(cancer|+,+)$, $P(\neg cancer|+,+)$

$$\Rightarrow P(cancer|+,+) = \frac{P(+,+|cancer)P(cancer)}{P(+,+)} = \frac{P(+|cancer)P(+|cancer)P(cancer)}{P(+,+)}$$

assume independent $\Rightarrow P(+,+|cancer) = P(+|cancer)P(+|cancer)$

$$\text{so } P(\neg cancer|+,+) = \frac{P(+,+|\neg cancer)P(\neg cancer)}{P(+,+)} = \frac{P(+|\neg cancer)P(+|\neg cancer)P(\neg cancer)}{P(+,+)}$$

$\because P(+|cancer)P(+|cancer)P(cancer) = 0.98 \times 0.98 \times 0.008 = 0.0076832$

$P(+|\neg cancer)P(+|\neg cancer)P(\neg cancer) = 0.03 \times 0.03 \times 0.992 = 0.0008928$

so: $P(cancer|+,+) = \frac{0.0076832}{0.008576} = 0.895896$ $\quad P(\neg cancer|+,+) = 0.104104$