2

a

(i)

In my simulation, the error rate is 0.2, not equals to 0.3.

The error rate may equal to 0.3 and it may not equal to 0.3 because it is random. And N=10 is not enough to predict actual error rate.

(ii)

If we want error rate equals to 0.3. In that draw, 3 data points are incorrect and 7 are correct. So,

$$ p = {10 } 0.3^{3} (1-0.3)^{7} = 0.27 $$

(iii)

If we want error rate equals to μ. In that draw, μN data points are incorrect and $(1-\mu)N$ are correct. So,

$$ p = {N N} ^{N} (1-)^{(1-) N}$$

b

(i)

max error rate = 0.7

min error rate = 0.0

sample mean of error rate = 0.31

sample std of error rate = 0.146

(ii)

72 runs had error rates different than μ

I think this agrees with the value of (a)(i)

(ii)

Probability is 0.28

c

μ	N	Theoretically $P(E(h) = \mu)$
0.1	10	0.387
0.1	100	0.131
0.3	10	0.27
0.3	100	0.0867
0.5	10	0.246
0.5	100	0.0795

μ	N	max	min	mean	std	# runs	P
0.1	10	0.3	0.1	0.096	0.084	54	0.46
0.1	100	0.19	0.02	0.099	0.028	85	0.9
0.3	10	0.7	0.0	0.31	0.146	72	0.28
0.3	100	0.47	0.18	0.3	0.048	95	0.81
0.5	10	1.0	0.2	0.54	0.149	73	0.27
0.5	100	0.64	0.38	0.49	0.048	91	0.68

d

(i)

Estimations based on N = 100 is more accurate then N = 10

(ii)

I don’t think the classifier learns something because the mean error rates are still around 0.5 and according to $P(|E(h) - \mu| < 0.05)$, the prediction is not so accurate.

For $\mu = 0.5$ N = 10, 76 datasets indicate the classifier learns something.

For $\mu = 0.5$ N = 100, 38 datasets indicate the classifier learns something.