1.

(a). $\mu=0.05$

$$P[k \mid N,\mu] = \binom{N}{k} \mu^k (1-\mu)^{N-k}$$

1 coin:

at least one coin has $v=0$.

$$P(0 \mid 10, 0.05) = \binom{10}{0} 0.05^0 (1-0.05)^{10} = 0.5987$$

1000 coins:

$$P[v \neq 0] = 1-0.5987 = 0.4013$$

$$P[\text{at least 1 coin will have } v=0] = 1-(0.4013)^{1000} \approx 1$$

1,000,000 coins:

$$P[\text{at least 1 coin will have } v=0] = 1-(0.4013)^{1{,}000{,}000} \approx 1$$

$\mu=0.8$.

1 coin:

$$P[0 \mid 10, 0.8] = \binom{10}{0} 0.8^0 (0.2)^{10} = 1.024\times10^{-7}$$

$$P[v \neq 0] = 1-1.024\times10^{-7}$$

1,000 coins.

$$P[\text{at least 1 coin will have } v=0] = 1-(1-1.024\times10^{-7})^{1000}$$

1,000,000 coins.

$$P[\text{at least 1 coin will have } v=0] = 1-(1-1.024\times10^{-7})^{1000000}$$

b. (i) 1000 coins are the data size N.

(ii) probability of the training error equals to 0.

(iii) $\mu = 0.05$

$M = 2^{10}$

$\nu = \frac{K}{10}$

$$\varepsilon(M, N, \delta) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}}$$

2. (a)

$$\varepsilon(1, N, 0.03) = \sqrt{\frac{1}{2N} \ln \frac{2 \times 1}{0.03}} \leq 0.05$$

$$\frac{1}{2N} \ln \frac{2}{0.03} \leq 0.05^2$$

$$\ln \frac{2}{0.03} \leq 0.05^2 \times 2N.$$

$$N \geq \frac{4.1997}{2 \times 0.05^2}$$

$$N \geq 839.94$$

$$N \geq 840$$

(b)

$$\varepsilon(1, N, 0.03) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \leq 0.05$$

$$\ln \frac{2 \times 100}{0.03} \leq (0.05)^2 \cdot 2N$$

$$N \geq \frac{8.8049}{2 \times 0.05^2}$$

$$N \geq 1761$$

$$N \geq 1761$$

(c)

$$\varepsilon(1, N, 0.03) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \leq 0.05$$

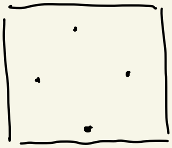
$$\ln \frac{2 \times 10^{k}}{0.03} \leq 0.05^2 \cdot 2N.$$

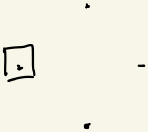
$$N \geq \frac{13.41}{2 \cdot 0.05^2}$$

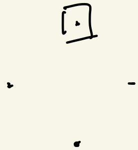
$$N \geq 2682$$

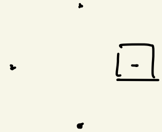
$$N \geq 2682$$

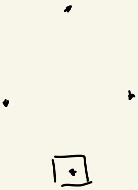
3.

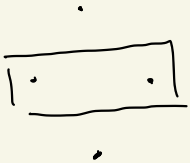

1

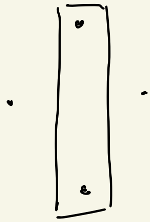

2

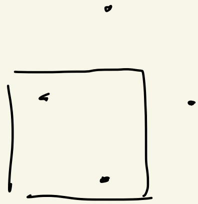

3

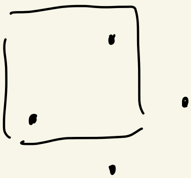

4

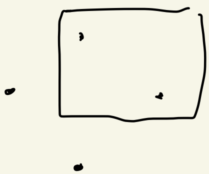

5

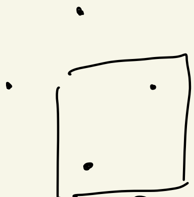

6

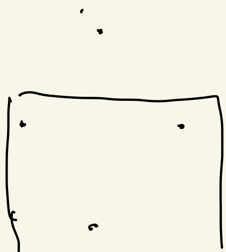

7

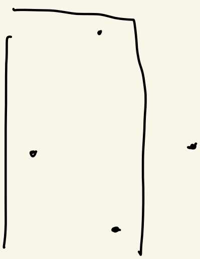

8

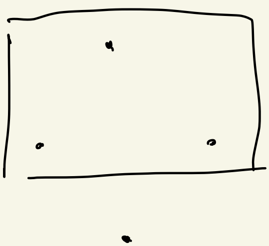

9

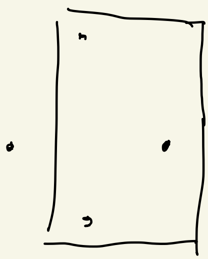

10

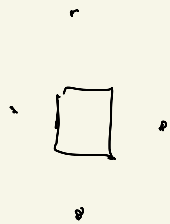

11


12

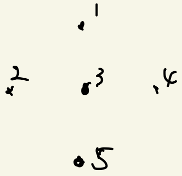

13


14


15


16

(b)



When there are 5 data points, a rectangle cannot put 1 and 5, or 2 and 4 in the same group.

$m_H(5) \leq 2^5$

(c) $d_{vc} = 4$

$m_H(N) \leq N^4 + 1$

4.

(a) $k=2$

$$M_H(2) = N+1 = 3 < 2^2$$

(b)

$$M_H(N) = \frac{1}{2}N^2 + \frac{1}{2}N + 1 = 2^N.$$

$$\frac{1}{2} \times 9 + \frac{1}{2} \times 3 + 1 =$$

$K=3$

$$M_H(3) = \frac{1}{2} \times 3^2 + \frac{1}{2} \times 3 + 1 = 7 < 2^3$$

5.

(a) $\varepsilon(1, N, 0.03) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} \leq 0.05$

$\varepsilon_{train}(1000, N, 0.05) = \sqrt{\frac{1}{2N} \ln \frac{2M}{\delta}} = \sqrt{\frac{1}{2 \cdot 400} \ln \frac{2 \cdot 1000}{0.05}} = 0.1151$

$\varepsilon_{test}(1, 200, 0.05) = \sqrt{\frac{1}{2 \cdot 200} \ln \frac{2}{0.05}} = 0.096$

Ein has a higher 'error bar'

(b) We want to have enough dat points to choose the best hypothesis g. Reserve more data points for test data. we may not have enough data points to choose the best hypothesis g.

6. (a) $\bar{g} = E_D[g^D(x)]$ $g^D(x) \in H$ H contains any linear combination. of hypotheses in H. so the expected function of all $g^D(x)$ is also in H. Therefore $\bar{g} \in H$

(b) Flip a coin 1000 times. Take heads as 1, and tail as 0. $H \in \{0, 1\}$ $\bar{g}$ should be close to ½, which is not in the hypotheses.

(c) No. $\bar{g}$ should be a number between 0 and 1.