Problem 1

d is the correct answer.
We have that

$$\epsilon = \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}(2N)}{\delta}}$$

We are using the simple approximate bound $N^{d_{VC}}$ for the growth function $m_{\mathcal{H}}(N)$. So this equation becomes

$$\epsilon = \sqrt{\frac{8}{N} \ln \frac{4(2N)^{10}}{\delta}}$$

We can now see that $\epsilon = .0507$ at $N = 440,000$ and $\epsilon = .0496$ at $N = 460,000$. So we have our answer.

Problem 2

d is the correct answer.
For the first two, I just plugged in all the numbers into the right side to get the bound. For the second two, I solved the equality for ϵ by plugging all the numbers in to get the bound. In doing so, I got the fourth bound was the smallest for $N = 10000$.

Problem 3

c is the correct answer.
For the first two, I just plugged in all the numbers into the right side to get the bound. For the second two, I solved the equality for ϵ by plugging all the numbers in to get the bound. For all this plug and chug, I used $N = 5$. This made d_{VC} irrelevant since the growth function always took in less than 50 points, so the growth function $m_{\mathcal{H}}(N)$ was always just 2^N. In doing so, I got the third bound was the smallest.

Problem 4

e is the correct answer.
For this problem I did it two ways. First, I ran linear regression, where the X matrix had the two x values and the Y matrix had the two y values of those x values. I ran the regression for 10,000 data sets and then took the average weight vector. The second way I did this problem was to just try to minimize the mean squared error for a random pair of points, for a large number of trials. So for each random pair of points in the space, I would try slopes from -5 to 5, in small increments, and pick the one that gave me the least mean squared error. I then took the average slope over all the trials. Both these methods gave me a slope of around 1.43, which is not an answer option.

Problem 5

b is the correct answer.
For this problem I ran a for loop from $x = -1$ to $x = 1$ in increments of .0005 and found the bias at each x. I then averaged the bias to get my answer.

Problem 6

a is the correct answer.
For this problem, I first calculated var(x) by trying 10000 data sets and finding the difference of the data set hypothesis slope and the average hypothesis slope (1.43) squared. I then averaged this value over all the data sets. I took this value and found its expected value over the interval by multiplying it by x^2 from $x = -1$ to $x = 1$ in increments of .0005, and finding the average of all those values.

Problem 7

b is the correct answer.
We have that the expected value for out-of-sample error for choice **a** is .75 from lecture. We also have that the expected value for out-of-sample error for choice **c** is 1.9 from lecture. And we have that the expected value for out-of-sample error for choice **b** is .5 from our answers to problems 5 and 6. Then we have that hypotheses that **d** and **e** represent are too complex given the data resources (and we can visualize how these choices would not do so well given 2 points to approximate the target function). So we have our answer.

Problem 8

c is the correct answer.
We have that

$$m_{\mathcal{H}}(N+1) = 2m_{\mathcal{H}}(N) - \binom{N}{q} = 2^{N+1}$$

as long as $\binom{N}{q} = 0$. And $\binom{N}{q} \geq 0$ when $N \geq q$. So we have our answer.

Problem 9

b is the correct answer.
When we take the intersection of all our hypothesis sets, we cannot possibly get more elements than are in the hypothesis set that has the minimum VC dimension. In other words, we are at least limited to those elements that are in the hypothesis set with the minimum VC dimension. So clearly

$$d_{VC}(\cap_{k=1}^{K}\mathcal{H}_k) \leq \min\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K}$$

Then we have that the VC dimnsion is at least 0. So we have our answer.

Problem 10

Choose two: **d** and **e**
When we take the union of all our hypothesis sets, we include all the elements that are in the hypothesis set that has the maximum VC dimension. So we have that

$$\max\{d_{VC}(\mathcal{H}_k)\}_{k=1}^{K} \leq d_{VC}(\cup_{k=1}^{K}\mathcal{H}_k)$$

Then we have that if an upper bound for the d_{VC} is $Y = \sum_{k=1}^{K} d_{VC}(\mathcal{H}_k)$,

$$\sum_{i=0}^{d_{VC}(\mathcal{H}_1)} \binom{N}{i} + \cdots + \sum_{i=0}^{d_{VC}(\mathcal{H}_k)} \binom{N}{i} < 2^{Y+1}$$

must be true. And if this inequality is true, then the d_{VC} is at most Y. To see how we reached this inequality, remember that $m_{\mathcal{H}_i}(N) \leq \sum_{i=0}^{d_{VC}(\mathcal{H}_i)} \binom{N}{i}$; that is, the most dichotomies on N points for a hypothesis set H_i is bounded above by that summation. Then we have that if we take the union of all the hypothesis sets, the maximum number of dichotomies the union will be able to get on N points is the sum of these bounds, because in the best case the dichotomies that each hypothesis set gets are independent from the dichotomies that the other hypothesis sets get. And we want the left side to be less than 2^{Y+1} because we want to show that $Y+1$ is a break point. Finally by visualizing this inequality with Pasqual's triangle, we can see that it is true (it helps to remember the property that $\sum_{i=0}^{N} \binom{N}{i} = 2^N$). So we have our answer. We will also put **e** just in case this reasoning doesn't hold, because **e** is the only other answer with the tightest lower bound.