Homework 6

Problem 1

b is the correct answer.
H' is less sophisticated as H, so it likely approximates the target function worse than H. Thus, we will have more deterministic noise.

Problem 2

a is the correct answer.
See attached code. Similar to hw 2, I ran linear regression after non-linear transformation. After getting weights from the in-sample points, I calculated the in-sample and out-of-sample classification errors to be 0.0285 and 0.084, respectively, closest to answer choice a.

Problem 3

d is the correct answer.
See attached code. Running weight decay with k = -3 (formula from slide 11 of lecture 12), we get the in-sample and out-of-sample classification errors to be 0.0285 and 0.08, respectively, closest to answer choice d.

Problem 4

e is the correct answer.
See attached code. Using the same code as problem 3, we change k to 3, and we get the in-sample and out-of-sample classification errors to be 0.371 and 0.436, respectively, closest to answer choice e.

Problem 5

d is the correct answer.
See attached code. I had an array of possible values for k, consisting of the answer choices and for each value of k, calculated out-of-sample classification error. At the end of the loop, we get the min. out of sample classification error and the k it belonged to. In this case, we get the minimum out of sample error to be 0.056, belonging to k = -1, or answer choice d.

Problem 6

b is the correct answer.
See attached code. This is very similar to problem 5, but instead of having values of k being limited to answer choices, we let the range of k go from -20 to 20 and keeping track of the minimum out of sample classification error. We get the minimum out of sample error to be 0.056, closest to answer choice b.

Problem 7

c is the correct answer.
The intersection of the two sets has it so that when $w_q = 0$ for $q \geq 3$. This is because H(10, 0, 3) is included in H(10, 0, 4). When $w_q = 0$ for $q \geq 3$, we get H_2, or answer choice c.

Problem 8

d is the correct answer.
For the forward propagation, we use

$$x_j^{(l)} = \theta(\sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)})$$

Thus, for each j we need $d^{i-1} + 1$ operations. For $l = 1$, we need 6 * 3 = 18 operations and for $l = 2$, we need 4 * 1 = 4 operations, so total, we have 18 + 4 = 22 operations.
For the backwards, we use the formula:

$$\delta_i^{(l-1)} = (1 - (x_i^{(l-1)})^2) \sum_{j=1}^{d^{(l)}} w_{ij}^{(l)} \delta_j^{(l)}$$

This means that for each $\delta_i^{(l-1)}$, we need d^l operations. We already know $\delta_i^{(2)}$ so we don't have to calculate that. We never use $\delta_i^{(0)}$ so there is no need to calculate that. Thus, we only need to calculate for $l = 2$. This means we need 3 * 1 = 3 operations. For updating weights, we use: $x_i^{(l-1)} \delta_j^{(l)}$, which are of the form

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta x_i^{(l-1)} \delta_j^{(l)}$$

Thus, for each w_{ij}, we have one operation. There are 6 * 3 + 4 * 1 = 22 of these. In total, we have 22 + 3 + 22 = 47 operations, closest to answer choice d.

Problem 9

a is the correct answer.
The minimum comes from having two units for each hidden layer. Thus, we have 10 weights going from the first layer to the non-constant node of the second layer and then 36 weights going from one hidden layer to the next, all the way up to the output. 10 + 36 = 46, so answer choice a.

Problem 10

e is the correct answer.
We try the case where there are just two hidden layers. To find the weights in this case, we have that w = 10 (h - 1) + x * (36 -h - 1) + (36 - h), where h is the number of nodes in the first hidden layer. Maximizing the above equation, we get that w = 510 when there are 22 nodes in the first hidden layer. 510 is the greatest of all the answer choices, so we get choice e.