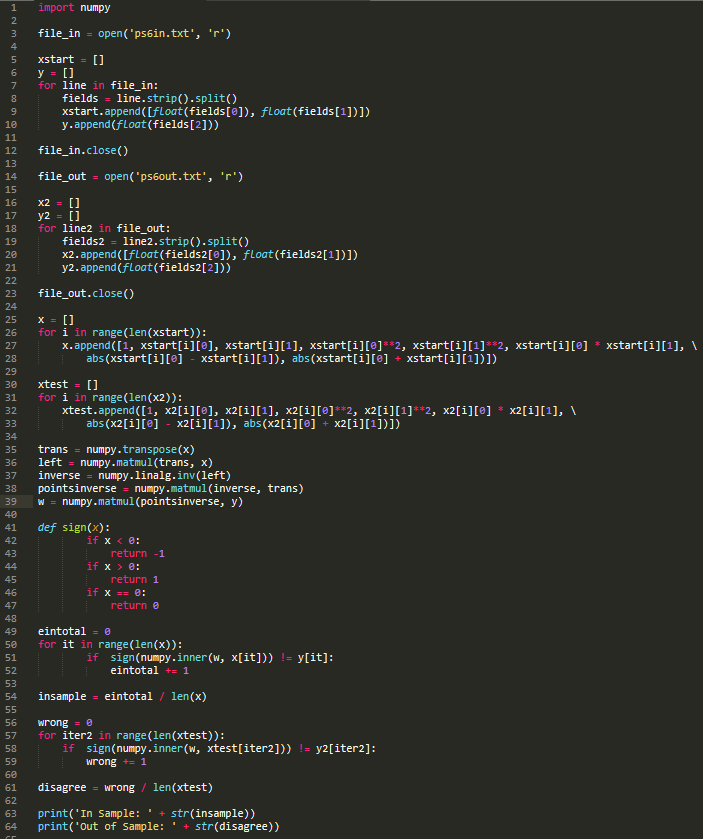
CS156a Problem Set #6

1. If we make the hypothesis set less complex while keeping the target function the same, the deterministic noise will increase. There will be more parts of target function that the hypothesis set fails to cover, so overfitting will go up. By decreasing the complexity of the hypothesis set, we are creating a bigger difference between the functions which then increases the noise of the hypothesis set. Although certain subsets can cause the bias to be as good or better, the noise will generally be bigger. The answer is [b].

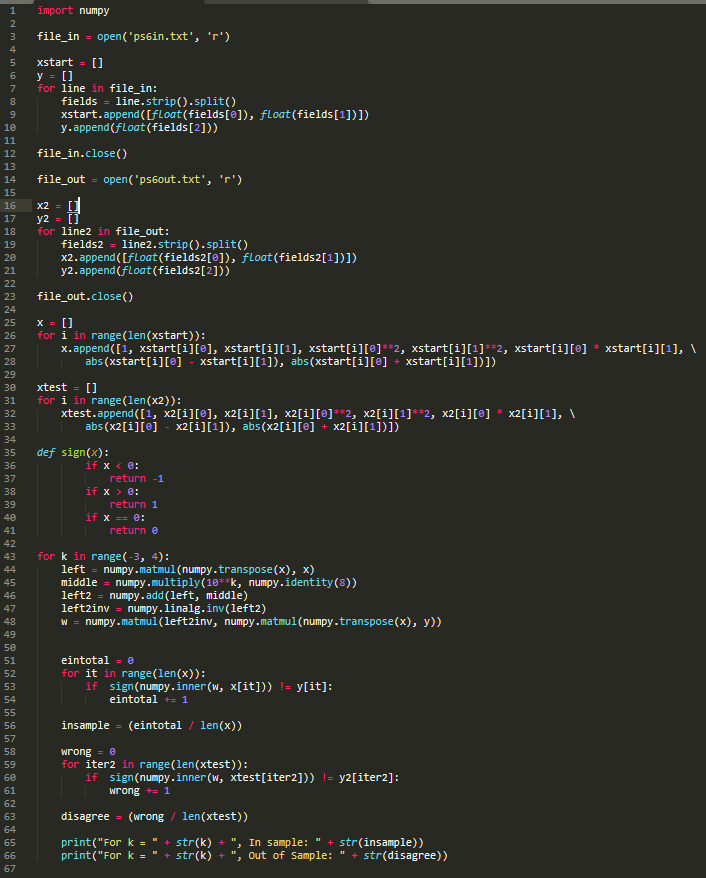
2.

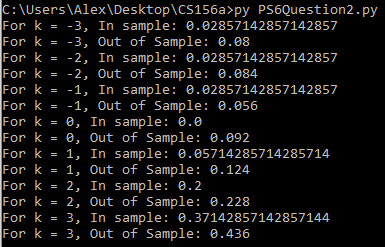




The answer is closest to [a].

3. Building off my old code:



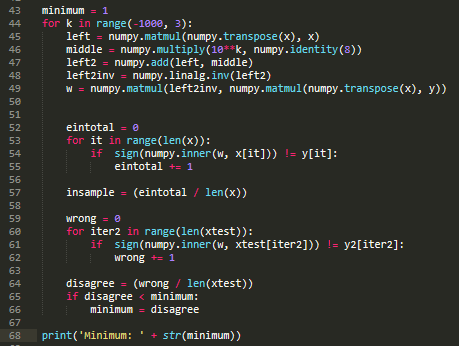


The answer is closest to [d].

4. The answer is closest to [e].

5.The answer with the lowest error is [d].

6. Since we know that the error decreases as k gets smaller (to a certain point), we can check through negative values to make sure. Changing just the last for loop in the previous code:





The answer is closest to [b].

7. We must consider each option:

a. The union of two the two sets will result in the union of H2 and H3 since all the terms for q = 3 and q = 4 respectively will turn to zero, since C = 0. Under the definition of our hypothesis set, this is not equivalent to H4. This is wrong.

b. The union of the two sets will be very large. It will union the two summations that will both be the sum of eleven terms (including q = 0), which is nowhere close to H3. This is wrong.

c. Both H(10, 0, 3) and H(10, 0, 4) will start adding zero at q = 3 and q = 4 respectively, so the values that will intersect that are nonzero will be those with q < 3. Looking at the summation representation of wTz, we will have a sum from q = 0 to q = 2, which is equal to H2. This is correct.

d. This is like c except now there is also an intersect of the points for q > 3, since the weights at those points will be one by the constraint. Now, we have a larger sum of points multiplied by weights for q < 3 and points for q > 3, which is not H1. This is wrong.

Thus, the answer is [c].

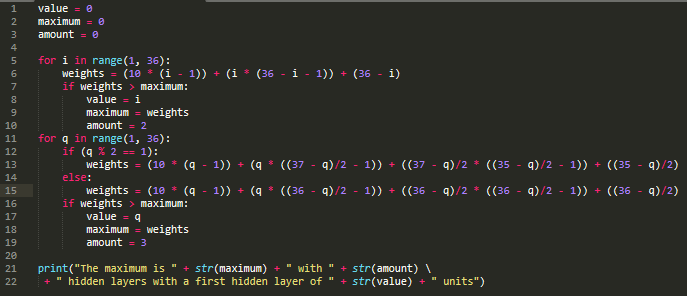
8. The backpropagation algorithm has three operations per iteration: forward computation of all xj(l), backward computation of all j(l), and updating the weights wij(l). We have L = 2, so we have to iterate from zero to two only focusing on the three products. With this information, we can calculate out the operations: 5 \* 3 + 3 \* 1 + 5 \* 3 + 3 \* 1 + 5 + 3 + 1 = 45 operations.

The answer is [d].

9. The minimum possible of weights can be made by having the ten inputs and then having 18 hidden layers before the output layer (maximizing the number of hidden layers). Thus, we have ten weights from the input layer to the first hidden layer, only having a weight to one value in the first hidden layer. Then, two weight between all the hidden layers and the last hidden layer to the output layer. This gives us 10 + 2 \* 18 = 46 weights.

The answer is [a].

10. We can find the maximum possible of weights by creating several hidden layers based on numbers that sum to 36. A general trend is the fewer hidden layers, the higher the weights. However, one hidden layer does not do much (386 weights). Thus, we can try different combinations of 2 and 3 hidden layers:





Thus, the maximum is when we have two hidden layers with 22 units in the first layer and 14 in the second. The answer is [e].