

CS 156 Problem Set 6

Christopher Zhen

Nov 7, 2016

Problem 1

(B) - Since we're trying to fit the same function with a smaller hypothesis set, our hypothesis generally won't fit as well, so there will be an increase in deterministic noise.

Problem 2

(A) - According to our code, our in sample error is 0.028 and our out of sample error is 0.084 which is closest to answer A.

Problem 3

(D) - Our code outputs errors similar to before (0.029 and 0.08) because lambda is very small and doesn't make much of a difference.

Problem 4

(E) - This time we get huge errors (0.37 and 0.44) due to the large lambda.

Problem 5

(D) - Even though $k = 0$ gives you the best in-sample error (error = 0), the best out-of-sample error is attained when $k = -1$ with an out-of-sample error of 0.056.

Problem 6

(B) - The code outputs an error of 0.056 which is closest to answer choice B when using $k = -1$. Something interesting to note is that if we didn't have the out-of-sample points, we would've chosen $k = 0$ which would've given us an out-of-sample error of 0.092 which is significantly worse.

Problem 7

(C) - In answer choice C, $\mathcal{H}(10, 0, 3)$ is equivalent to \mathcal{H}_2 and $\mathcal{H}(10, 0, 4)$ is equivalent to \mathcal{H}_3 . The intersection of \mathcal{H}_2 and \mathcal{H}_3 is simply \mathcal{H}_2 , so answer choice C is correct.

Problem 8

(D) - First we need to compute the gradient which is done by summing all the $w_{ij}^{(l)} x_i^{(l-1)}$ terms. This comes out to be $6 * 3 + 4 * 1 = 22$. Next we need to compute all the required $\delta_j^{(l)}$ values. Since we only need these partials for three nodes in the middle layer (since we already know it for the output node and we don't need it for $x_0^{(1)}$), we only need $1 * 3 = 3$ computations to find all the $\delta_j^{(l)}$ values. Finally, we need to update the weights which is done by multiplying all the $w_{ij}^{(l)}$ by the corresponding $\delta_j^{(l)}$ terms. This number also comes out to 22 since each $w_{ij}^{(l)}$ maps to a single $\delta_j^{(l)}$, so we just count the $w_{ij}^{(l)}$ terms again. Summing up we get $22 + 3 + 22 = 47$ which is closest to answer choice D.

Problem 9

(A) - The neural network with the least amount of weights involves the least number of nodes in each layer (2) to minimize multiplicative terms. This gives us 10 weights going into the first hidden layer and then 36 weights until the output node; or 46 total weights.

Problem 10

(E) - In this case we want to have 2 hidden layers because having more hidden layers just adds more terms at the cost of multiplying more terms which would give us fewer weights. With two hidden layers, we have the following expression that we need to maximize (where x is the number of nodes in the first hidden layer):

$$10(x - 1) + x(35 - x) + 36 - x \tag{1}$$

This expression is maximized when $x = 22$ with the value 510 which matches answer choice E.