Learning from Data
Homework # 6

Khoi Pham

Overfitting and Deterministic Noise

1. Deterministic noise represents the inability of a hypothesis set to approximate the target function. Thus, if we are to use a smaller subset of hypotheses $\mathcal{H}'$ rather than the full hypothesis set $\mathcal{H}$, then it becomes more difficult to approximate the target function. In other words, the deterministic noise increases in this case.
Answer: [b]

Regularization with Weight Decay

The code for answering questions in this section is uploaded here.

2. As the result in the code has shown, the in-sample error is 0.0285 and the out-of-sample error is 0.084.
Answer: [a]

3. Using $\lambda = 0.001$, the in-sample error is still 0.0285, but the out-of-sample error reduces to 0.08.
Answer: [d]

4. Using $\lambda = 1000$, the in-sample error is 0.0371, and the out-of-sample error is 0.436.
Answer: [e]

5. According to the code, the best value of k is -1, and the best out-of-sample error is 0.056.
Answer: [d]

6. Best out-of-sample error is 0.056 achieved with $k = -1$.
Answer: [b]

Regularization for Polynomials

7. Answer **c** is correct because $\boldsymbol{\mathcal{H}_2}$ contains up to second-order polynomial, and it is correctly represented using option **c**.
Answer: [c]

Neural Networks

8. Both the $w_{ij}^{(l)} x_i^{(l-1)}$ and $x_i^{(l-1)} \delta_j^{(l)}$ computations take $(d^{(0)}+1) \times d^{(1)} + (d^{(1)+1} \times d^{(2)})$ operations. And the $w_{ij}^{(l)} \delta_j^{(l)}$ computation takes $d^{(1)} \times d^{(2)}$ operations. So the total number of operations is $22 + 22 + 3 = 47$.
Answer: [d]

9. In order to get the minimum number of weights of the network, we can construct it by making all of the hidden layers contain only one node each. The number of weights will be $10 + 1 \times 36 = 46$.
Answer: [a]

10. Suppose we are constructing a network with 2 hidden layers with x_1 and x_2 nodes each respectively. The number of weights are

$$\begin{aligned} & 10 \times (x_1 - 1) + x_1(x_2 - 1) + x_2 \\ = & \, 10 \times x_1 - 10 + x_1 x_2 - x_1 + x_2 \end{aligned}$$

Because $x_2 = 36 - x_1$, we substitute this into x_2 to get

$$\begin{aligned} & 10 \times x_1 - 10 + 36x_1 - x_1^2 - x_1 + 36 - x_1 \\ = & -x_1^2 + 44x_1 + 26 \end{aligned}$$

Differentiate the above equation and set it to 0, we find that $x_1 = 22, x_2 = 14$ produces the largest number of weights, which is $-22^2 + 44 \times 22 + 26 = 510$.
Answer: [e]