(7) [C]

HW6

The hypothesis set $\mathcal{H}(Q, C, Q_0)$ where $C=0$ would create a set of polynomials of order $\min(Q, Q_0-1)$ orless (as $\mathcal{H}_n$ is order n or less)

as seen below:

$$\sum_{q=0}^{Q-1} w_q L_q(x) + \sum_{q=Q_0}^{Q} 0 L_q(x)$$

because $w_q = C = 0$ when $q \geq Q_0$, as seen above, then we use an upper bound of $Q_0 - 1$. However Q could also be a potential upper bound for when $C \neq 0$;

- $\mathcal{H}(10,0,3) = \{h \mid h(x) = w^T z \in \mathcal{H}_Q = \sum_{q=0}^{10} w_q L_q(x),\ w_q = 0 \text{ for } q \geq 3\} = \{h \mid h(x) = \sum_{q=0}^{2} w_q L_q(x)\} = \mathcal{H}_2$
- $\mathcal{H}(10,0,4) = \{h \mid h(x) = w^T z \in \mathcal{H}_Q = \sum_{q=0}^{10} w_q L_q(x),\ w_q = 0 \text{ for } q \geq 4\} = \{h \mid h(x) = \sum_{q=0}^{3} w_q L_q(x)\} = \mathcal{H}_3$
- $\mathcal{H}(10,1,3) = \{h \mid h(x) = w^T z \in \mathcal{H}_Q = \sum_{q=0}^{10} w_q L_q(x),\ w_q = 1 \text{ for } q \geq 3\} = \{h \mid h(x) = \sum_{q=0}^{2} w_q L_q(x) + \sum_{q=3}^{10} L_q(x)\} = \mathcal{H}_{10}$
- $\mathcal{H}(10,1,4) = \{h \mid h(x) = w^T z \in \mathcal{H}_Q = \sum_{q=0}^{10} w_q L_q(x),\ w_q = 1 \text{ for } q \geq 4\} = \{h \mid h(x) = \sum_{q=0}^{3} w_q L_q(x) + \sum_{q=4}^{10} L_q(x)\} = \mathcal{H}_{10}$

Now:

[a] $\mathcal{H}(10,0,3) \cup \mathcal{H}(10,0,4) = \mathcal{H}_4$

a is false because:

$\mathcal{H}(10,0,3) = \mathcal{H}_2$ & $\mathcal{H}(10,0,4) = \mathcal{H}_3$ as shown above hence their union is $\mathcal{H}_3$ not $\mathcal{H}_4$.

[b] $\mathcal{H}(10,1,3) \cup \mathcal{H}(10,1,4) = \mathcal{H}_3$

b is false because:

$\mathcal{H}(10,1,3) = \mathcal{H}_{10}$ & $\mathcal{H}(10,1,4) = \mathcal{H}_{10}$ as shown above hence their union is greater than order 3 hence it's not $\mathcal{H}_3$ as it is for order 3 polynomials and less.

[c] $\mathcal{H}(10,0,3) \cap \mathcal{H}(10,0,4) = \mathcal{H}_2$

C is correct because:

$\mathcal{H}(10,0,3) = \mathcal{H}_2$ & $\mathcal{H}(10,0,4) = \mathcal{H}_3$ as shown above hence their intersection is $\mathcal{H}_2 \cap \mathcal{H}_3 = \mathcal{H}_2$

[d] $\mathcal{H}(10,1,3) \cap \mathcal{H}(10,1,4) = \mathcal{H}_1$

d is false because:

$\mathcal{H}(10,1,3) = \mathcal{H}_{10}$ & $\mathcal{H}(10,1,4) = \mathcal{H}_{10}$ as shown above hence their intersection is greater than $\mathcal{H}_1$.

(8) [d]

$1 \le l \le 2$ ← layers; $0 \le i \le 5$ inputs $l=1$; $1 \le j \le 3$ outputs $l=1$

$0 \le i \le 3$ inputs $l=2$; $1 \le j \le 1$ outputs $l=2$

• Forward compute all $x_j^{(l)}$ for $L=2$:

$$x_j^{(l)} = \theta\left(\sum_{i=0}^{d^{(l-1)}} w_{ij}^{(l)} x_i^{(l-1)}\right)$$ as seen in lecture 10 slide 14

$$x_1^{(1)} = \theta\left(\sum_{i=0}^{5} w_{ij}^{(1)} x_i^{(0)}\right)$$

$$x_1^{(1)} = \theta\left(w_{01}^{(1)} x_0^{(0)} + w_{11}^{(1)} x_1^{(0)} + w_{21}^{(1)} x_2^{(0)} + w_{31}^{(1)} x_3^{(0)} + w_{41}^{(1)} x_4^{(0)} + w_{51}^{(1)} x_5^{(0)}\right)$$

$$x_2^{(1)} = \theta\left(w_{02}^{(1)} x_0^{(0)} + w_{12}^{(1)} x_1^{(0)} + w_{22}^{(1)} x_2^{(0)} + w_{32}^{(1)} x_3^{(0)} + w_{42}^{(1)} x_4^{(0)} + w_{52}^{(1)} x_5^{(0)}\right)$$

$$x_3^{(1)} = \theta\left(w_{03}^{(1)} x_0^{(0)} + w_{13}^{(1)} x_1^{(0)} + w_{23}^{(1)} x_2^{(0)} + w_{33}^{(1)} x_3^{(0)} + w_{43}^{(1)} x_4^{(0)} + w_{53}^{(1)} x_5^{(0)}\right)$$

$$x_1^{(2)} = \theta\left(\sum_{i=0}^{3} w_{i1}^{(2)} x_i^{(1)}\right) = \theta\left(w_{01}^{(2)} x_0^{(1)} + w_{11}^{(2)} x_1^{(1)} + w_{21}^{(2)} x_2^{(1)} + w_{31}^{(2)} x_3^{(1)}\right)$$

• Now count all $w_{ij}^{(l)} x_i^{(l-1)}$ we get $(3 \times 6) + 4 = 18 + 4 = 22$ operations

Backwards compute all:

$$\delta_i^{(l-1)} = \left(1 - (x_i^{(l-1)})^2\right) \sum_{j=1}^{d^{(l)}} w_{ij}^{(l)} \delta_j^{(l)}$$ from lecture slide 14.

(let this be k) — we only care about this part as it is in the desired operation form.

for $l=1$:

$$\delta_i^0 = k \sum_{j=1}^{3} w_{ij}^{(1)} \delta_j^{(1)}$$

Since $l-1=0$ then the number of operations for the input layer is zero as there is no signal to any node.

for $l=2$:

$$\delta_1^1 = k\left(w_{01}^{(2)} \delta_1^{(2)}\right)$$ ← we will have 2 more of these operations for $i=2,3$ hence $3 \times 1 = 3$ total operations

• Now for updating the weights:

$$w_{ij}^{(l)} = w_{ij}^{(l)} + \eta x_i^{(l-1)} \delta_j^{(l)}$$ from lecture slide 20. (we only care about this part)

for $l=1$; $0 \le i \le 5$; $1 \le j \le 3$

$w_{01}^{(1)} = w_{01}^{(1)} \eta x_0^{(0)} \delta_1^{(1)}$ ← we will have a total of $3 \times 6 = 18$ of these operations for $j=1,2,3$ & $i=0,1,2,3,4,5$ as for each j we must iterate through all i's.

for $l=2$: $0 \le i \le 3$; $1 \le j \le 1$

$w_{01}^{(2)} = w_{01}^{(2)} + \eta x_0^{(1)} \delta_1^{(2)}$ Again we will have a total of $4 \times 1 = 4$ of these operations for $j=1$ & $i=0,1,2,3$ as for each j we must iterate through all i's.

Hence we have a total number of operation equal to $4 + 18 = 22$ for weight updates.

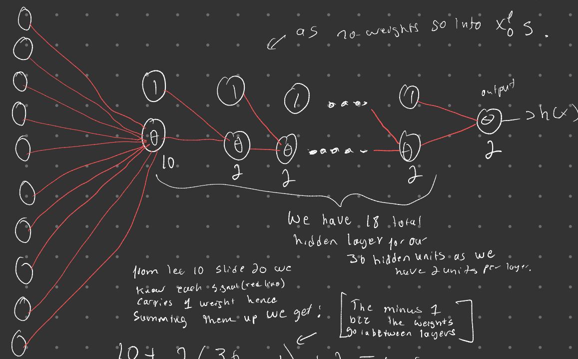
Hence our total number of all 3 operations is:

$22 + 3 + 22 = 47$ hence the correct answer is [d] as it is the closest.

(9) [a]

The minimum possible number of weights is achieved when we put 2 units in each hidden layers as seen below and as similarly drawn in lec 10 slide 13.

← as no weights go into x_0^l s.



We have 18 total hidden layer for our 36 hidden units as we have 2 units per layer.

from lec 10 slide 20 we know each signal (red line) carries 1 weight hence summing them up we get:

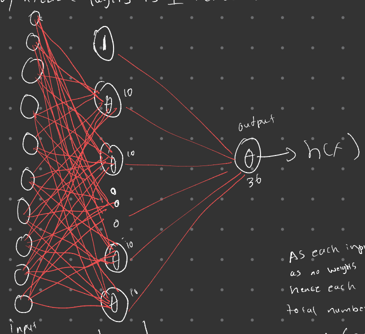
[The minus 1 b/c the weights go in between layers]

$$10 + 2\left(\frac{36}{2} - 1\right) + 2 = 46 \text{ weights}$$

hence the correct answer is [a]

(10) [e]

We will achieve the maximum number of possible weights when we minimize our hidden layers. lets look at when the number of hidden layers is 1 hence similar to the intuition in 9 we will have:



As each input makes 35 connections as no weights go into x_0^1 hence each has 35 weights then the total number of weights are:

$10(35) + 36 = 386$ ← maximum for 1.

• Now for the case we have 2 hidden layers with 18 units in each then as seen by the previous networks, all 10 inputs connect to 17 units in the first hidden layer as x_0s dont receive any weight. Then each of the 18 units in the first hidden layer will connect to 17 units in hidden layer 2. Then each of those 18 units will connect to our one output hence giving us:

$(10 \times 17) + (18 \times 17) + 18 = 494$ connections hence 494 weights. This also gives us a general formula for total number of weights for 2 layers $= 10(x-1) + x(36 - x - 1) + (36 - x)$ for x being the number of units for the first layer as the units are connected.

$w = 10x - 1 + 36x - x^2 - x + 36 - x$

$w' = 10 + 36 - 2x - 1 - 1 = -2x + 44$; Find critical points

$0 = 44 - 2x$; $x = 22$; (+ | − at 22) hence 22 is a maximum

∴ we have a max w when $x = 22$, hence 22 units in layer 1.

for $x = 22 \rightarrow w(22) = 10(22-1) + 22(36-22-1) + (36-22) = 510$ hence is the maximum weight.

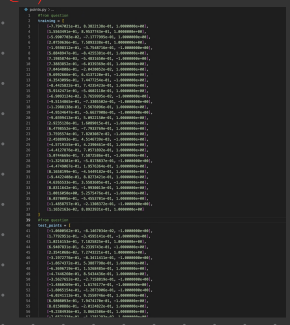
Hence the correct answer is [e]

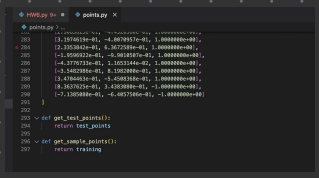
(1) [b]

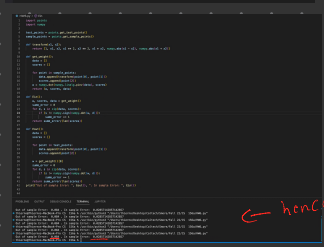
Deterministic noise is our bias as seen in lecture 11 slide 20. As we know from lecture 8 slide 15, our Bias increases with less complex hypothesis sets hence increasing deterministic noise for H' as it is a subset of H hence it is less complex.

By this reasoning the correct answer is [b].

(2) [a]

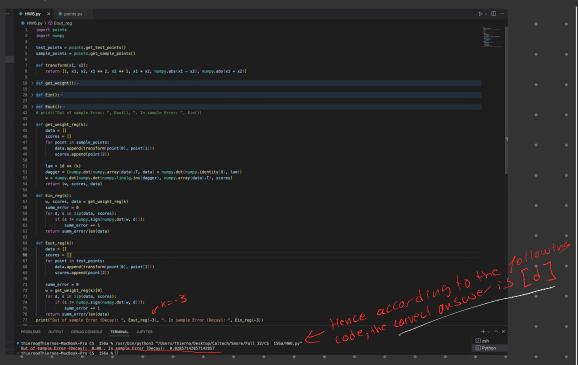






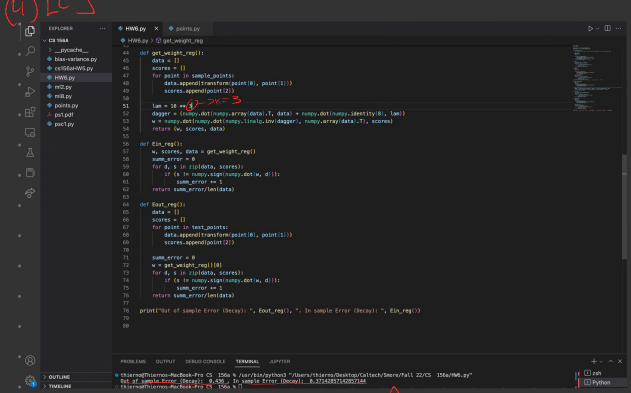
← hence according to the code the correct answer is [a]

(3) [d]



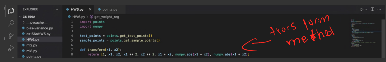
← Hence according to the following code, the correct answer is [d]

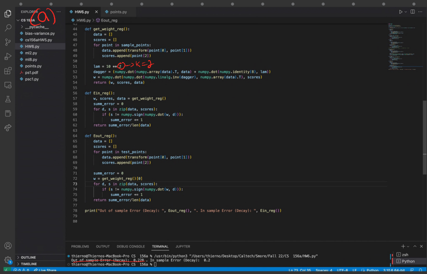
(4) [c]

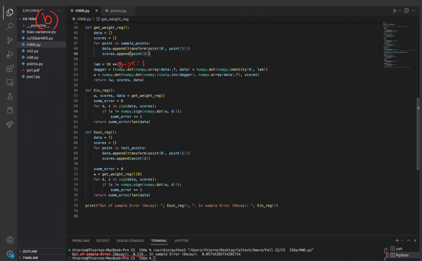


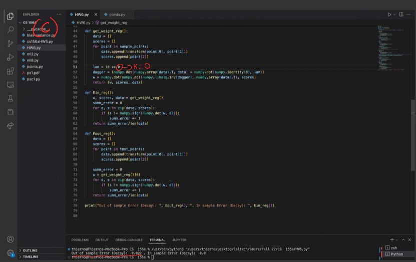
Hence according to the following code, the correct answer is [c] as it is closest.

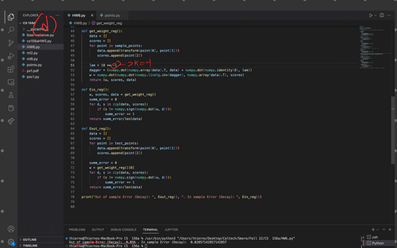
(5) [d]

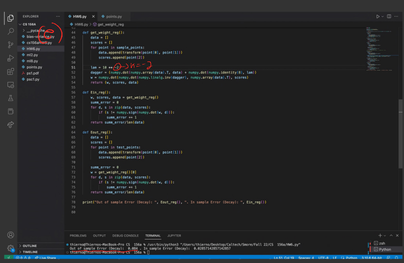






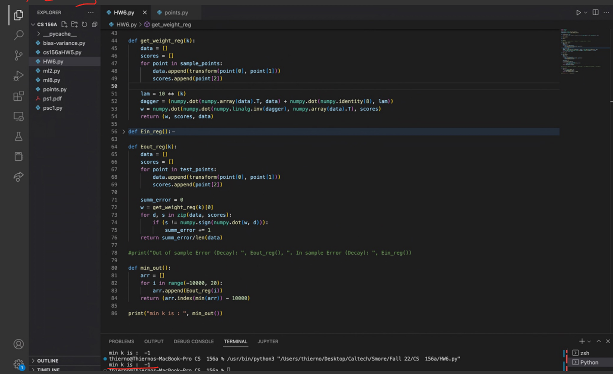


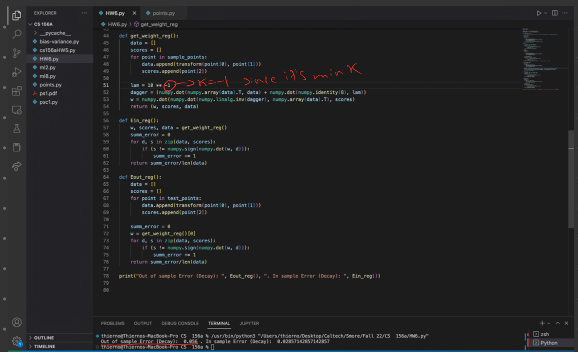




Hence according to the above code correct answer is [d]

(6) [b]





Hence by the above code the correct answer is [b] as it is the closest.