

$$8.1 (a) \quad 12 + \min(12, 3) + 3 = 12 + 3 + 3 = 18$$

Where 12 come from $3 \times 4 = 12$

$$(b) = 3 + 4 + 4 = 11$$

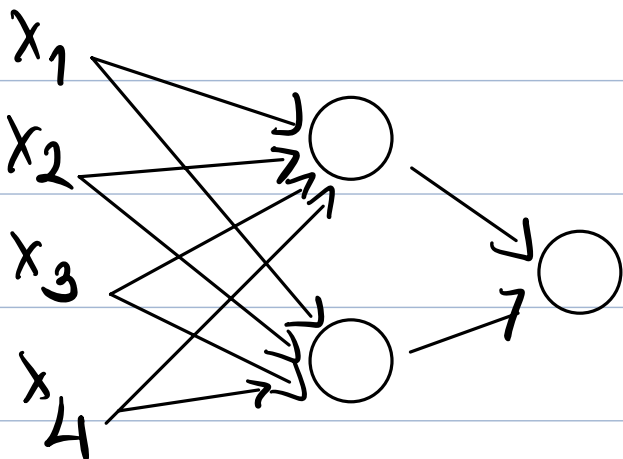
(c) For the first one we have 18 and second one we have 11

(d) we have that for first network $\frac{18}{\log_2(4)} = 9$

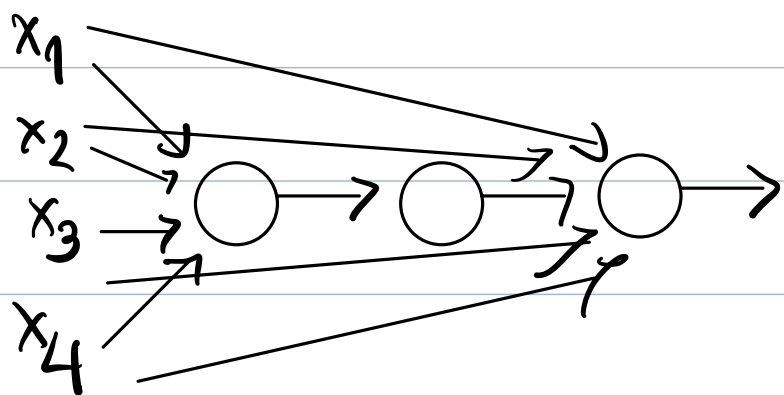
and second network $\frac{11}{\log_2(4)} = 5$

8.2

First network



second network



8.4

(c) We have that the visual information we have is that we assume that we are active 16 hours per day and our eye can process about 1 megapixel and each pixel is represented by 24 bits then assume that our life span is 80 years then we have that $(10^6 \times 24) \times 60 \times 60 \times 16 \times 365 \times 80$
 $= 4.03608 \times 10^{16}$ bits. For Shakespeare we have that there are about 5,000,000 words and 10 bits per word then it is around 50 million bits. For our brain capacity we have that with 100 billion neurons and each making 1,000 connections with 2 bits then $10^{11} \times 10^3 \times 2 = 2 \times 10^{14}$ bits. However, we also have that with a little bit of compression of the data then the calculation kind of shows us that we would be able to remember almost everything.

(b) We have that to expand Algorithms to then the algorithm need to account for multiple class labels not just two. which therefore mean that it instead of counting transitions from 0 to 1 and vice versa it would need to count transitions between any 2 different classes as well. Additionally, we can store the thresholds_matrix that store the thresholds for each pair of classes. Also for the MEC calculation for multiclass we may need to consider the worst case scenario which is that the max number of transitions from one class to any other class. We also have that we can show empirically so that the answer will move to be $\frac{C}{C-1}$ where C is number of class

C6) For regression we have that we might not need to count the thresholds or classes. However, instead we can use the average of the labels or regression values for each unique sum of features as our predicted value for any type of other regressions and use the number of rows of table for the thresholds to

calculate the MEC. Additionally, we can think of regression as a class classification so the $\frac{C}{C-1}$ we have that in this case our regression will just be number of points is a lot then this will tend to 1.



