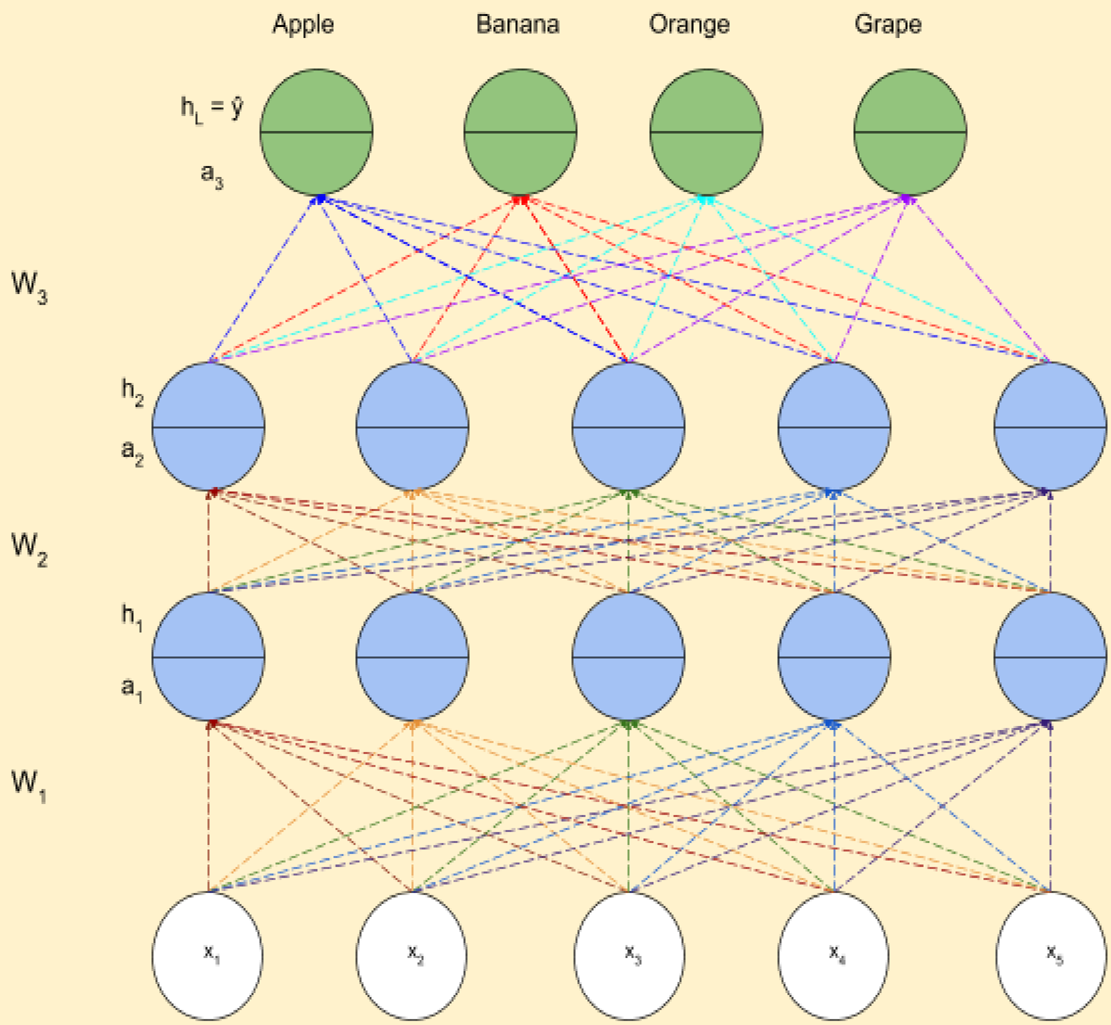
PadhAI: Deep Neural Networks

One Fourth Labs

Output Layer of a Multi-Class Classification Problem

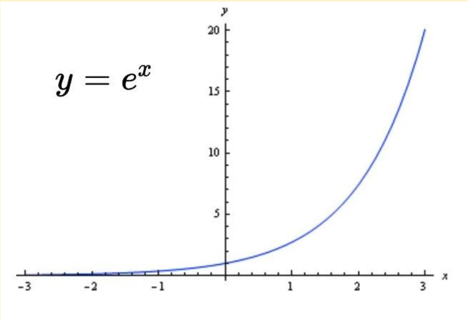
Deciding the output layer

1. The Output Activation function is chosen depending on the task at hand (can be a softmax, linear etc)
2. Consider the following multi-class classification problem



3. At the last layer, we compute $a_3 = W_3 h_2 + b_3$
4. We need to apply a function to $\hat{y}_i = O(a_{3i})$ such that the 4 outputs form a probability distribution.
5. Output activation function has to be chosen such that the output is probability.
6. Let's assume $a_3 = [3 \;\; 4 \;\; 10 \;\; 3]$
   a. Take each entry and divide by the sum of all entries to get a probability distribution
   b. $\hat{y}_1 = \frac{3}{(3+4+10+3)} = 0.15$
   c. $\hat{y}_2 = \frac{4}{(3+4+10+3)} = 0.20$
   d. $\hat{y}_3 = \frac{10}{(3+4+10+3)} = 0.50$
   e. $\hat{y}_4 = \frac{3}{(3+4+10+3)} = 0.15$
   f. However, **this will not work** if any of the entries are negative

7. So we consider the softmax function
8. $softmax(z_i) = \frac{e^{z_i}}{\Sigma_{j=1}^{k} e^{z_j}}$ for i = 1...k
9. Note: the output of e^x is always positive, irrespective of the input



10. This property is important to counter the negative-value shortcoming that we observed in the previous example
11. Now, let us illustrate the softmax function at the last layer of our Neural Network
12. $a = [a_1\ a_2\ a_3\ a_4]$
13. $softmax(a) = [\frac{e^{a_1}}{\Sigma_{j=1}^{k} e^{a_j}}\ \frac{e^{a_2}}{\Sigma_{j=1}^{k} e^{a_j}}\ \frac{e^{a_3}}{\Sigma_{j=1}^{k} e^{a_j}}\ \frac{e^{a_4}}{\Sigma_{j=1}^{k} e^{a_j}}]$
14. Raising the numerators to e^x ensures that they are all positive
15. The denominator is just the sum of all the values raised to e^x
16. $softmax(a_i)$ *is the* i^{th} *element of the softmax output*
17. So for our multi-class fruit classifier, the equations are as follows

Layer	**Pre-activation**	**Activation/Output**
Hidden Layer 1	$a_1 = W_1 * x + b_1$	$h_1 = g(a_1)$
Hidden Layer 2	$a_2 = W_2 * h_1 + b_2$	$h_2 = g(a_2)$
Output Layer	$a_3 = W_3 * h_2 + b_3$	$\hat{y} = softmax(a_3)$