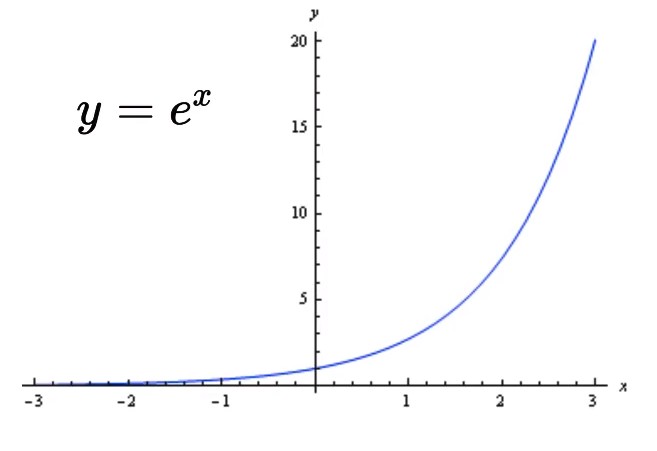
### **Output Layer of a Multi-Class Classification Problem**

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 a3 = W3h2 + b3
4. We need to apply a function to ŷi = O(a3i) 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 a3 = [3 4 10 3]
   1. Take each entry and divide by the sum of all entries to get a probability distribution
   2. However, **this will not work** if any of the entries are negative
7. So we consider the softmax function
8. for i = 1...k
9. Note: the output of ex 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. Raising the numerators to ex ensures that they are all positive
13. The denominator is just the sum of all the values raised to ex
14. So for our multi-class fruit classifier, the equations are as follows

|  |  |  |
| --- | --- | --- |
| **Layer** | **Pre-activation** | **Activation/Output** |
| Hidden Layer 1 |  |  |
| Hidden Layer 2 |  |  |
| Output Layer |  |  |