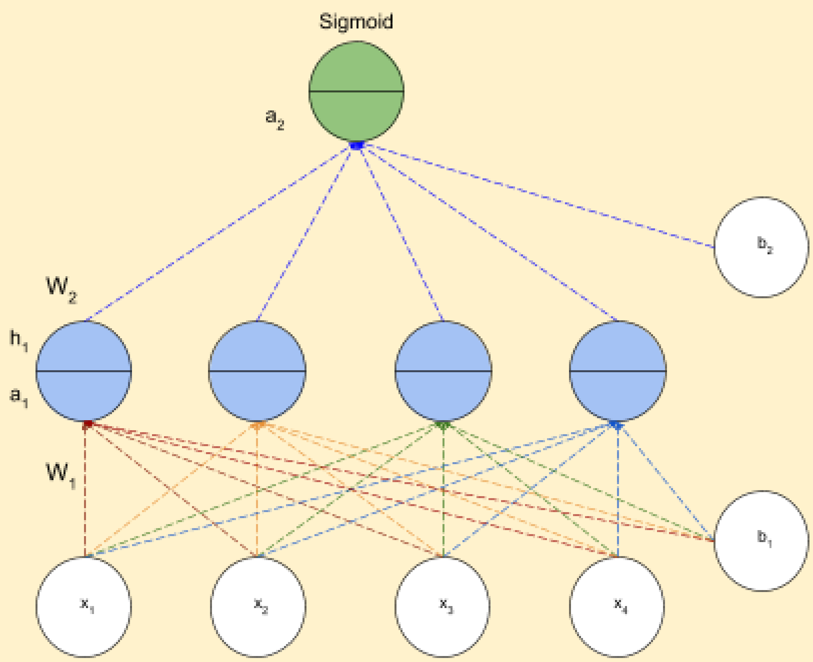
Loss function

Loss function for Binary Classification

What is the loss function that you can use for a binary classification problem

1. In normal cases, the number of neurons in the output layer would be equal to the number of classes
2. However a shortcut in the case of binary classification would be to use only one output neuron that uses a sigmoid function. Here is a diagrammatic representation of that configuration



3. Here, $\hat{y} = P(y = 1)$
4. Therefore, we can obtain $P(y = 0) = 1 - P(y = 1)$
5. Consider the following values for the variables
   a. $b = [\, 0.5 \;\; 0.3 \,]$
   b. $y = 1$

$$W_1 = \begin{bmatrix} 0.9 & 0.2 & 0.4 & 0.3 \\ -0.5 & 0.4 & 0.3 & 0.3 \\ 0.1 & 0.1 & -0.1 & 0.2 \\ -0.2 & 0.5 & 0.5 & / \end{bmatrix}$$

   c. $W_2 = [0.5 \;\; 0.8 \;\; -0.6 \;\; 0.3]$
   d. $x = [0.3 \;\; 0.5 \;\; -0.4 \;\; 0.3]$

6. The output values are as follows
   a. $a_1 = W_1 * x + b_1 \quad = [0.8 \ \ 0.52 \ \ 0.68 \ \ 0.67]$
   b. $h_1 = sigmoid(a_1) \quad = [0.69 \ \ 0.63 \ \ 0.66 \ \ 0.67]$
   c. $a_2 = W_2 * h_1 + b_2 \quad = 0.948$
   d. $\hat{y} = sigmoid(a_2) = 0.7207$
   e. In this case $y = 1 \quad True\ distribution\ [0 \ \ 1]$
   f. $Predicted\ distribution\ \hat{y} \quad [0.2793 \quad 0.7207]$
   g. Cross Entropy Loss:
      i. $L(\Theta) = (y)(-\log(\hat{y})) + (1-y)(-\log(\hat{y}))$
      ii. In this case, since y = 1
      iii. $L(\Theta) = -1 * log(0.7207)$
      iv. $L(\Theta) = 0.327$
7. Consider another case where x = [-0.6 -0.6 0.2 0.3] and true class y = 1
8. The output values are as follows
   a. $a_1 = W_1 * x + b_1 \quad = [0.01 \ \ 0.71 \ \ 0.42 \ \ 0.63]$
   b. $h_1 = sigmoid(a_1) \quad = [0.50 \ \ 0.67 \ \ 0.60 \ \ 0.65]$
   c. $a_2 = W_2 * h_1 + b_2 \quad = 0.921$
   d. $\hat{y} = sigmoid(a_2) = 0.7152$
   e. In this case $y = 0 \quad True\ distribution\ [1 \ \ 0]$
   f. $Predicted\ distribution\ \hat{y} \quad [0.2848 \quad 0.7152]$
   g. Cross Entropy Loss:
      i. $L(\Theta) = (y)(-\log(\hat{y})) + (1-y)(-\log(\hat{y}))$
      ii. In this case, since y = 0
      iii. $L(\Theta) = -1 * log(1 - 0.7152)$
      iv. $L(\Theta) = 1.2560$
      v. Here, even though the true value was 0, our neuron was outputting a very large value(0.7152) which was already indicative of a large loss value.