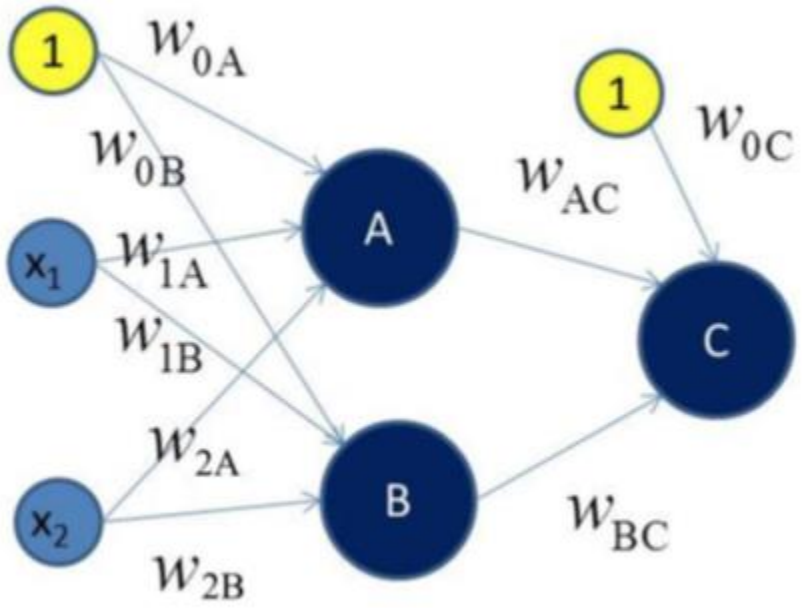
1. Feed forward the following MLP neural network given the following: Weights: all weights are 0.1 Activation: both the hidden layer and the output layer use a sigmoid activation Default value of the bias inputs (in yellow): 1 Input data (x1=0.5, x2=0.8).



Screenshot from Homework assignment (Prof. Lauria, N.D.).

To feedforward an MLP Neural Network means that the network is bound to a unidirectional flow and does not allow for looping or cycling which means that for this kind of neural network, we are moving from left to right direction and, once we reach at an output, the network is complete.

Given data:

All weights are 0.1, X_1 =0.5, X_2 = 0.8, Since all of our weights are 0.1,

W_{0A}= W_{0B}= W_{1A}= W_{1B}= W_{2A}= W_{2B}= W_{0C}= W_{AC}= W_{BC}

netA = 1 * W_{0A} + X_1 * W_{1A} + X_2 * W_{2A}

= (1 * 0.1) + (0.5 * 0.1) + (0.8 * 0.1)

netA = 0.23

netB = 1 * W_{0B} + X_1 * W_{1B} + X_2 * W_{2B}

= (1 * 0.1) + (0.5 * 0.1) + (0.8 * 0.1)

netB = 0.23

The 0.23 value we got is used as an input to our sigmoid activation function. The equation using our new input is:

aA = sigmoid (netA) = 1 / (1 + $e^{(-0.23)}$) = 0.557

aB = sigmoid (netB) = 1 / (1 + $e^{(-0.23)}$) = 0.557

netC = 1 * W_{0C} + aA * W_{AC} + aB * W_{BC}

= (1 * 0.1) + (0.557 * 0.1) + (0.557 * 0.1)

netC = 0.2114

yC = sigmoid (netC) = 1/ (1 + $e^{(-0.2114)}$) = 0.5527

The output of our neural network is 0.5527 on the first pass.

2. Should we prefer a large hidden layer (many nodes) or a small one? Describe the benefits and drawbacks of each.

One of the drawbacks of a large hidden layer is that too many nodes overfitting the model and lowering testing accuracy over a smaller one. Having many nodes in a hidden layer allows model to capture the complexity in a dataset but a we need to check on how the model performs and understand the chances of overfitting. On the other hand, having too few hidden layer nodes could create a model with low training accuracy thus underfitting the data. It really seems that there is no particular answer as such and it totally depends on the experience in working or understating the subject. To have the best output, always use optimum layers which neither underfit nor overfit the data.