159740, Studies in Intelligent Systems

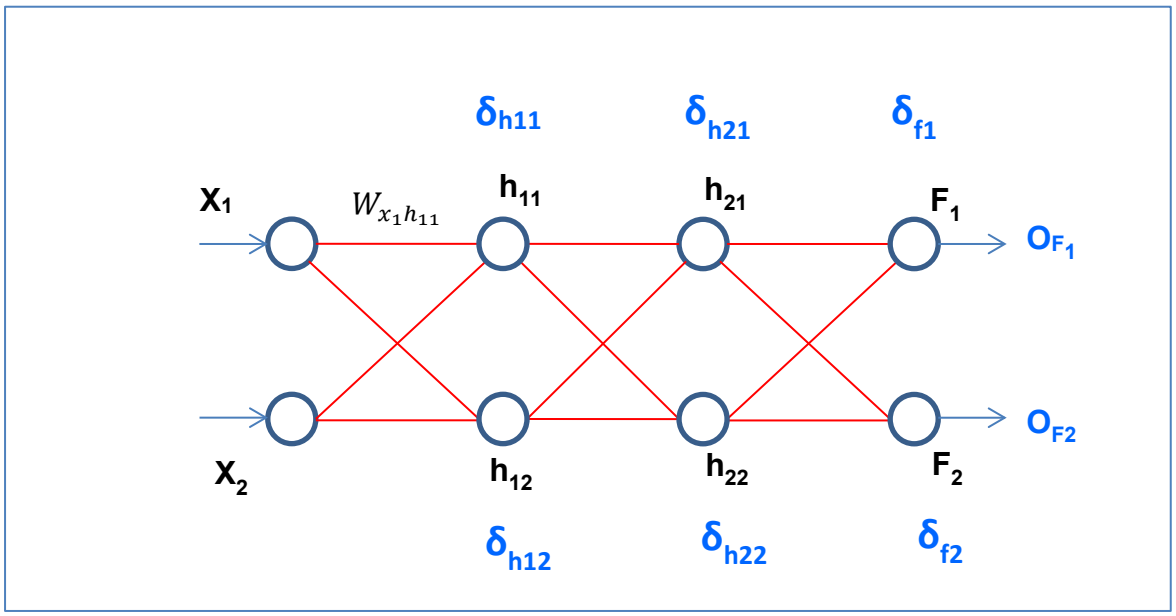
Tutorial #6: Derivation of Back-propagation Formulas for a 4-layer Neural Network (with Softmax units and ReLU)

Instructions

Given the deep neural network architecture below, write all the formulas required for updating the weights in the network. Use Gradient Descent, on-line learning method.

Neural network architecture:

- 2 inputs
- 2 outputs
- 2 hidden layers
- Activation function for output nodes: **Softmax**
- Activation function for all hidden nodes= **ReLU**



Formula for calculating the output of the softmax units

1. $o_{F_1} = \dfrac{\exp(netInput_{F_1})}{\exp(netInput_{F_1}) + \exp(netInput_{F_2})}$
2. $o_{F_2} = \dfrac{\exp(netInput_{F_2})}{\exp(netInput_{F_1}) + \exp(netInput_{F_2})}$

Updating Weights leading into the output nodes

3. $\delta_{F_1} = (o_{F_1} - d_{F_1})$
4. $W_{h_{21}F_1} = W_{h_{21}F_1} - \eta\delta_{F_1} o_{h_{21}}$
5. $W_{h_{22}F_1} = W_{h_{22}F_1} - \eta\delta_{F_1} o_{h_{22}}$
6. $\delta_{F_2} = (o_{F_2} - d_{F_2})$
7. $W_{h_{21}F_2} = W_{h_{21}F_2} - \eta\delta_{F_2} o_{h_{21}}$

8. $W_{h_{22}F_2} = W_{h_{22}F_2} - \eta\delta_{F_2} o_{h_{22}}$

Updating weights leading into the nodes at hidden layer 2

9. /// $\delta_{h_{21}} = o_{h_{21}}(1-o_{h_{21}}) \times \sum_{j=1}^{m} \delta_{F_j} W_{h_{21}F_j}$

If (h_{21} <= 0)

$\partial_{h21} = 0$

Else

$\partial_{h21} = \sum_{j=1}^{m} \left(\partial_{F_j} * W_{h_{21}F_j}\right)$

End If

10. $W_{h_{11}h_{21}} = W_{h_{11}h_{21}} - \eta\delta_{h_{21}} o_{h_{11}}$
11. $W_{h_{12}h_{21}} = W_{h_{12}h_{21}} - \eta\delta_{h_{21}} o_{h_{12}}$
12. /// $\delta_{h_{22}} = o_{h_{22}}(1-o_{h_{22}}) \times \sum_{j=1}^{m} \delta_{F_j} W_{h_{22}F_j}$

If (h_{22} <= 0)

$\partial_{h22} = 0$

Else

$\partial_{h22} = \sum_{j=1}^{m} \left(\partial_{F_j} * W_{h_{22}F_j}\right)$

End If

13. $W_{h_{11}h_{22}} = W_{h_{11}h_{22}} - \eta\delta_{h_{22}} o_{h_{11}}$
14. $W_{h_{12}h_{22}} = W_{h_{12}h_{22}} - \eta\delta_{h_{22}} o_{h_{12}}$

Updating weights leading into the nodes at hidden layer 1

15. // $\delta_{h_{11}} = o_{h_{11}}(1-o_{h_{11}}) \times \sum_{j=21}^{22} \delta_{h_j} W_{h_{11}h_j}$

If (h_{11} <= 0)

$\partial_{h11} = 0$

Else

$\partial_{h11} = \sum_{j=21}^{22} \left(\partial_{h_j} * W_{h_{11}h_j}\right)$

End If

16. $W_{x_1 h_{11}} = W_{x_1 h_{11}} - \eta \delta_{h_{11}} x_1$
17. $W_{x_2 h_{11}} = W_{x_2 h_{11}} - \eta \delta_{h_{11}} x_2$
18. // $\delta_{h_{12}} = o_{h_{12}}(1 - o_{h_{12}}) \times \sum_{j=21}^{22} \delta_{h_j} W_{h_{12} h_j}$

If (h_{12} <= 0)

$\partial_{h12} = 0$

Else

$\partial_{h12} = \sum_{j=21}^{22} \left(\partial_{h_j} * W_{h_{12} h_j} \right)$

End If

19. $W_{x_1 h_{12}} = W_{x_1 h_{12}} - \eta \delta_{h_{12}} x_1$
20. $W_{x_2 h_{12}} = W_{x_2 h_{12}} - \eta \delta_{h_{12}} x_2$