

Forward Pass and Backpropagation Summary in matrix

One hidden layer

Activation function : sigmoid $f()$

Input: $X_{1 \times N}$ N inputs

Weights_input_to_hidden: $W_{i_h}_{N \times M}$

Weights_hidden_to_output: $W_{h_o}_{M \times K}$

$Y_{1 \times K}$

Forward pass:

Hidden_layer_input $_{1 \times M} = X_{1 \times N} \times W_{i_h}_{N \times M}$

Hidden_layer_output $_{1 \times M} = f(\text{Hidden_layer_input})_{1 \times M}$

Output_layer_input $_{1 \times K} = \text{Hidden_layer_output}_{1 \times M} \times W_{h_o}_{M \times K}$

Output_y $_{1 \times K} = f(\text{output_layer_in})_{1 \times K}$

Backpropagation

Error = $(Y - \hat{Y})_{1 \times K}$

output_error_term:

$\delta_{1 \times K} = \text{Error} * f'(\text{output_layer_input}_{1 \times K})$ (* element wise product)

Hidden_error_term :

$\delta_{1 \times M} = \delta_{1 \times K} \times W_{h_o}_{K \times M} * f'(\text{hidden_layer_input}_{1 \times M})$

Delta_w_h_o $_{M \times K} = \text{learnrate} * \text{Hidden_layer_output}_{M \times 1} \times \delta_{1 \times K}$

Delta_w_i_h $_{N \times M} = \text{learnrate} * X_{N \times 1} \times \delta_{1 \times M}$