Considerthefollowingnetworkstructure.Youcanassumetheinitialweights.Assumebias
to bezeroforeasiercomputations.Giventhat < x1, x2, ˆy1, ˆy2 >=< 1, 1, 0, 1 > where ˆy is
the target.Assume = 0.9 and □ = 0.01.
drow the mlp and solve below
(a) Computetheforwardpropagationandgeneratetheoutput.UseReluforhiddenlayers
and Sigmoidactivationfunctionforoutputlayer.[2]
(b) ComputetheSoftmaxlossfunctionforbothoutputs.[1]
(c) Let theinitialweightsthatassumedbetheweights[attime(t-1).Computetheweights
v21, w12 and w22 at time t using SGD.[1.5]
(d) Lettheweightattime t betheonescomputedinpart(c).Computetheweights v21,
w12 and w22 at (t + 1) when momentumisused.[1.5]

ChatGPT

Given the input features x_1 and x_2 with initial weights and biases assumed as follows:

- $w_{11} = 0.2$, $w_{12} = 0.3$, $w_{21} = 0.4$, $w_{22} = 0.5$
- $v_{11} = 0.1$, $v_{12} = 0.2$, $v_{21} = 0.3$, $v_{22} = 0.4$
- $b_1 = 0$, $b_2 = 0$

Given input: $x_1 = 1$, $x_2 = 1$, target output: $\hat{y}_1 = 0$, $\hat{y}_2 = 1$

Learning rate: $\alpha = 0.01$, momentum term: $\beta = 0.9$

(a) Forward Propagation:

- The ReLU activation function is applied to the hidden layer, and the sigmoid activation function is applied to the output layer.

Forward propagation equations:
$z_1 = w_{11}x_1 + w_{21}x_2 + b_1$
$z_2 = w_{12}x_1 + w_{22}x_2 + b_1$
$h_1 = \mathrm{ReLU}(z_1)$
$h_2 = \mathrm{ReLU}(z_2)$
$o_1 = \sigma(v_{11}h_1 + v_{21}h_2 + b_2)$
$o_2 = \sigma(v_{12}h_1 + v_{22}h_2 + b_2)$

Let's calculate the forward propagation.

(b) Softmax Loss Function:

- The softmax loss function computes the loss based on the predicted output probabilities and the target labels.

Softmax loss function equation for each output:
$L = -\sum_i y_i \log(\hat{y}_i)$

Let's compute the softmax loss function for both outputs.

(c) Updating Weights using Stochastic Gradient Descent (SGD):

- Weights are updated using the gradient descent algorithm with the learning rate.

Weight update equation for each weight:
$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \alpha \frac{\partial L}{\partial w_{ij}}$

Let's compute the updated weights v_{21}, w_{12}, and w_{22} using SGD.

(d) Updating Weights with Momentum:

- Weights are updated using momentum, which incorporates the previous weight update direction.

Weight update equation with momentum for each weight:
$\Delta w_{ij}^{(t)} = \beta \Delta w_{ij}^{(t-1)} + (1-\beta)\frac{\partial L}{\partial w_{ij}}$
$w_{ij}^{(t+1)} = w_{ij}^{(t)} - \alpha \Delta w_{ij}^{(t)}$

Let's compute the updated weights v_{21}, w_{12}, and w_{22} using momentum.