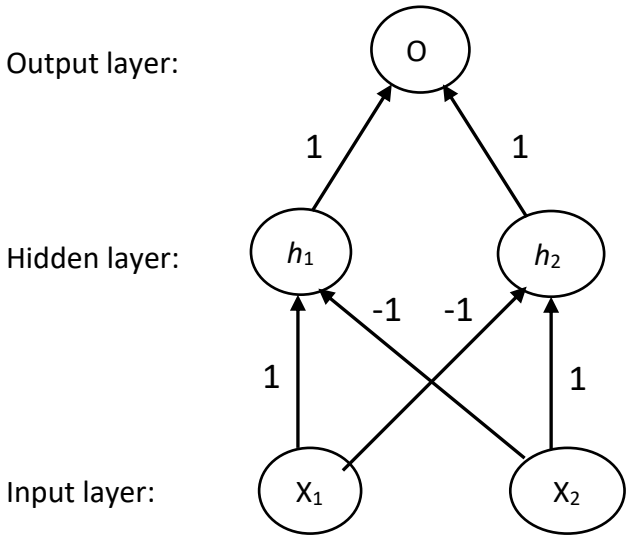

repeat
  for each e **in** *examples* **do**
    for each node j in the input layer **do** $a_j \leftarrow x_j[e]$
    for $\ell = 2$ **to** M **do**
      $in_i \leftarrow \sum_j W_{j,i}\ a_j$
      $a_i \leftarrow g(in_i)$
    for each node i in the output layer **do**
      $\Delta_i \leftarrow g'(in_i) \times (y_i[e] - a_i)$
    for $\ell = M - 1$ **to** 1 **do**
      for each node j in layer ℓ **do**
        $\Delta_j \leftarrow g'(in_j) \sum_i W_{j,i}\ \Delta_i$
        for each node i in layer $\ell + 1$ **do**
          $W_{j,i} \leftarrow W_{j,i} + \alpha \times a_j \times \Delta_i$
until some stopping criterion is satisfied

Sigmoid activation function
Learning rate 0.7
Trace the algorithm.
https://www.tinkershop.net/ml/sigmoid_calculator.html

XOR FFNet

X1	X2	y	h_1	h_2	h'_1	h'_2	O	O'	Δo	Δh1	Δh2	w'_{h1-O}	w'_{h2-O}
0	**0**	0	0.5	0.5	0.25	0.25	0.731	0.197	-0.144	-0.036	-0.036	-0.05	-0.05
0	**1**	1	0.269	0.731	0.197	0.197	0.721	0.201	0.056	0.010	0.010	0.011	0.029
1	**0**	1	0.731	0.269	0.197	0.197	0.724	0.2	0.055	0.01	0.011	0.028	0.01
1	**1**	0	0.504	0.504	0.250	0.250	0.730	0.197	-0.144	-0.036	-0.036	-0.051	-0.051

X1	X2	w_{h1-O}	w_{h2-O}	w'_{x1-h1}	w'_{x1-h2}	w_{x1-h1}	w_{x1-h2}	w'_{x2-h1}	w'_{x2-h2}	w_{x2-h1}	w_{x2-h2}
0	**0**	0.950	0.950	-0	-0	1	-1	-0	-0	-1	1
0	**1**	0.960	0.978	0	0	1	-1	0.007	0.007	-0.993	1.007
1	**0**	0.988	0.989	0.007	0.007	1.007	-0.993	0	0	-0.993	1.007
1	**1**	0.938	0.938	-0.025	-0.025	0.982	-1.017	-0.025	-0.025	-1.018	0.982