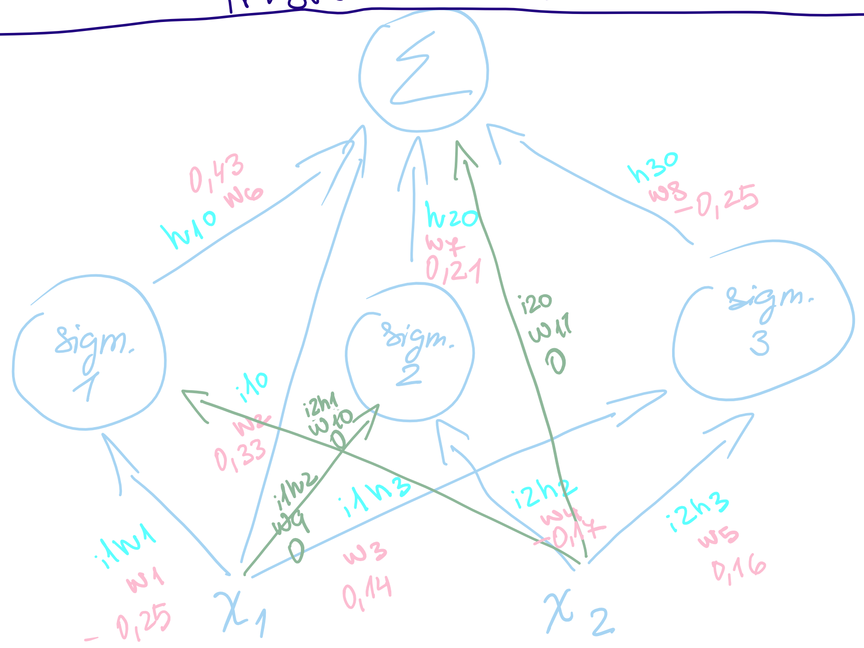
HAND CALCULATIONS FOR THE PROTOTYPE

ID: 33870918

Simple NN as outlined in the instructions



HAND CALCULATIONS (for the network above)

/. Network setup and forward pass

/. Weights and inputs:

/. Input to hidden weights:

$w_1 = -0{,}25$
$w_2 = 0{,}33$
$w_3 = 0{,}14$
$w_4 = -0{,}17$
$w_5 = 0{,}16$
$w_6 = 0{,}43$
$w_7 = 0{,}21$
$w_8 = -0{,}25$

$w_9 = 0$
$w_{10} = 0$
$w_{11} = 0$

/. Input vector:

$$x = \begin{bmatrix} 0{,}5 \\ -0{,}5 \end{bmatrix}$$

target(t) = 0,1

/. Hidden layer calculation

For each hidden neuron j, the net input is:

$$net_j = \sum_{i=1}^{2} w_{ij}^{(h)} x_i \quad (\text{bias} = 0)$$

=> 3 hidden neurons

Neuron 1: $-0{,}25 \cdot 0{,}5 + 0 \cdot (-0{,}5) = \underline{\underline{-0{,}125}}$

Neuron 2: $0{,}5 \cdot 0 + (-0{,}5) \cdot (-0{,}17) = \underline{\underline{0{,}085}}$

Neuron 3: $0{,}5 \cdot 0{,}14 + (-0{,}5) \cdot 0{,}16 = \underline{\underline{-0{,}01}}$

calculated as:
Neuron 1 $= x_1 \cdot w_1 + x_2 \cdot w_{10}$
Neuron 2 $= x_1 \cdot w_9 + x_2 \cdot w_4$
Neuron 3 $= x_1 \cdot w_3 + x_2 \cdot w_5$

/. Applying the sigmoid activation

$$h_j = \frac{1}{1+e^{-net_j}}$$

$$h_1 = \frac{1}{1+e^{0{,}125}} \approx \underline{\underline{0{,}4688}}$$

$$h_2 = \frac{1}{1+e^{-0{,}085}} \approx \underline{\underline{0{,}5212}}$$

$h_3 = \frac{1}{1+e^{0,01}} \approx 0,4975$

%. Output calculation

=> Hidden to output path:

$$y(h) = \sum_{j=1}^{3} w_j^{(0)} h_j$$

$$y(h) = w_6 \cdot h_1 + w_7 \cdot h_2 + w_8 \cdot h_3$$

$$y(h) = 0,43 \cdot 0,4688 + 0,21 \cdot 0,5212 + (-0,25) \cdot 0,4975$$

$$y(h) \approx \underline{\underline{0,1867}}$$

=> Direct input to output path:

$$y(d) = w_2^{(d)} x_1 + w_{11}^{(d)} x_2$$

$$y(d) = 0,33 \cdot 0,5 + 0 \cdot (-0,5)$$

$$y(d) = \underline{\underline{0,1650}}$$

=> Output:

$$y = y(h) + y(d)$$

$$y = 0,1867 + 0,1650 \approx \underline{\underline{0,3517}}$$

%. Error calculation

$$e = t - y = 0,1 - 0,3517 = \underline{\underline{-0,2517}}$$

%.%. Jacobian calculation

%. Input to hidden weights

derivative for a weight feeding into hidden neuron:

$$\frac{\partial y}{\partial w_{ij}^{(h)}} = w_j^{(0)} f'(net_j)\, x_i \quad , \text{with } f'(net_j) = h_j(1-h_j)$$

=> Hidden Neuron 1:

$\boxed{h_1 \approx 0,4688}$

$$f'(net_1) \approx h_1 \cdot (1-h_1)$$

$$f'(net_1) \approx 0,4688 \,(1-0,4688)$$

$$f'(net_1) \approx 0,4688 \cdot 0,5312$$

$$f'(net_1) \approx \underline{\underline{0,2490}}$$

for input 1: $w_{1,1}^{(h)}$

$$J_1 = w_6 \cdot f'(net_1) \cdot x_1$$

$$J_1 \approx 0,43 \cdot 0,2490 \cdot 0,5$$

$$J_1 \approx \underline{\underline{0,0535}}$$

for input 2: $w_{2,1}^{(h)}$

$$J_2 = w_6 \cdot f'(net_1) \cdot x_2$$

$$J_2 \approx 0,43 \cdot 0,2490 \cdot (-0,5)$$

$$J_2 \approx \underline{\underline{-0,0535}}$$

=> Hidden neuron 2:

$\boxed{h_2 = 0,5212}$

$$f'(net_2) \approx h_2 \cdot (1-h_2)$$

$$f'(net_2) \approx 0,5212 \cdot (1-0,5212)$$

$$f'(net_2) \approx \underline{\underline{0,2496}}$$

for input 1: $w^{(h)}_{1,2}$

$J_3 = w_7 \cdot f'(net_2) \cdot x_1$

$J_3 \approx 0{,}21 \cdot 0{,}2496 \cdot 0{,}5$

$J_3 \approx \underline{\underline{0{,}0262}}$

for input 2: $w^{(h)}_{2,2}$

$J_4 = w_7 \cdot f'(net_2) \cdot x_2$

$J_4 \approx 0{,}21 \cdot 0{,}2496 \cdot (-0{,}5)$

$J_4 \approx \underline{\underline{-0{,}0262}}$

=> Hidden neuron 3:

$\boxed{h_3 = 0{,}4975}$

$f'(net_3) = h_3 \cdot (1-h_3)$

$f'(net_3) \approx 0{,}4975 \cdot (1-0{,}4975)$

$f'(net_3) \approx 0{,}2499$

for input 1: $w^{(h)}_{1,3}$

$J_5 = w_8 \cdot f'(net_3) \cdot x_1$

$J_5 \approx -0{,}25 \cdot 0{,}2499 \cdot 0{,}5$

$J_5 \approx \underline{\underline{-0{,}0312}}$

for input 2: $w^{(h)}_{2,3}$

$J_6 = w_8 \cdot f'(net_3) \cdot x_2$

$J_6 \approx -0{,}25 \cdot 0{,}2499 \cdot (-0{,}5)$

$J_6 \approx \underline{\underline{0{,}0312}}$

% Hidden to output weights

$$\frac{\partial y}{\partial w_j^{(o)}} = h_j$$

$J_7 = h_1$	$J_8 = h_2$	$J_9 = h_3$
$J_7 \approx \underline{\underline{0{,}4688}}$	$J_8 \approx \underline{\underline{0{,}5212}}$	$J_9 \approx \underline{\underline{0{,}4975}}$

% Direct input to output weights

$$\frac{\partial y}{\partial w_i^{(d)}} = x_i$$

$J_{10} = x_1$	$J_{11} = x_2$
$J_{10} = \underline{\underline{0{,}5}}$	$J_{11} = \underline{\underline{-0{,}5}}$

Complete jacobian

$$J \approx [0{,}0535, -0{,}0535, 0{,}0262, -0{,}0262, 0{,}0312, -0{,}0312, 0{,}4688, 0{,}5212, 0{,}4975, 0{,}5, -0{,}5]$$

%% Gradient and Hessian approximation

$\boxed{t - y = e}$

$$E = \frac{1}{2}(e)^2 = \frac{1}{2}(-0{,}2517)^2$$

$E = \underline{\underline{0{,}0317}}$

$\nabla E = -J^T(-0{,}2517)$

$\nabla E = 0{,}2517\, J^T$

So, each component of the gradient is:

$(\nabla E)_k = 0{,}2517\, J_k$

Note: for a single training example ($N=1$), the expressions simplify as follows:

1) Cost function:
$E = \frac{1}{2}(t-y)^2$; $e = t-y$.

2) Gradient:
$\nabla E = -J^T e$, with $J = \left[\frac{\partial y}{\partial w_1}, \frac{\partial y}{\partial w_2}, \dots, \frac{\partial y}{\partial w_n}\right]$.

3) Gauss-Newton Hessian approximation:
$H \approx J^T J$.

4) Damped Hessian (Levenberg-Marquardt):
$H_{damped} = J^T J + \lambda I$.

SINCE $N=1$, THERE IS NO AVERAGING NEEDED, AND THE EXPRESSIONS REMAIN IN THIS FORMAT. DIVISION BY N IS TRIVIAL.

=> Calculating the gradient components:

$(\nabla E_1) = (\nabla E)_k \cdot J_1 \approx 0{,}2517 \cdot 0{,}0535 \approx \underline{0{,}0135}$

$(\nabla E_2) = (\nabla E)_k \cdot J_2 \approx 0{,}2517 \cdot (-0{,}0535) \approx -0{,}0135$

$(\nabla E_3) = (\nabla E)_k \cdot J_3 \approx 0{,}2517 \cdot 0{,}0262 \approx 0{,}0066$

$(\nabla E_4) = (\nabla E)_k \cdot J_4 \approx 0{,}2517 \cdot (-0{,}0262) \approx -0{,}0066$

$(\nabla E_5) = (\nabla E)_k \cdot J_5 \approx 0{,}2517 \cdot (-0{,}0312) \approx -0{,}0079$

$(\nabla E_6) = (\nabla E)_k \cdot J_6 \approx 0{,}2517 \cdot 0{,}0312 \approx 0{,}0079$

$(\nabla E_7) = (\nabla E)_k \cdot J_7 \approx 0{,}2517 \cdot 0{,}4688 \approx 0{,}1180$

$(\nabla E_8) = (\nabla E)_k \cdot J_8 \approx 0{,}2517 \cdot 0{,}5212 \approx 0{,}1312$

$(\nabla E_9) = (\nabla E)_k \cdot J_9 \approx 0{,}2517 \cdot 0{,}4975 \approx 0{,}1252$

$(\nabla E_{10}) = (\nabla E)_k \cdot J_{10} \approx 0{,}2517 \cdot 0{,}5 \approx 0{,}1259$

$(\nabla E_{11}) = (\nabla E)_k \cdot J_{11} \approx 0{,}2517 \cdot (-0{,}5) \approx -0{,}1259$

% Hessian approximation

$H \approx J^T J$

$H_{1,1} = 0{,}0535^2 \approx 0{,}0029$

$H_{1,2} = (-0{,}0535)^2 \approx 0{,}0029$

$H_{2,1} = 0{,}0262^2 \approx 0{,}0007$

$H_{2,2} = (-0{,}0262)^2 \approx 0{,}0007$

$H_{3,1} = 0{,}0312^2 \approx 0{,}0010$

$H_{3,2} = (-0{,}0312)^2 \approx 0{,}0010$

$H_4 = 0{,}4688^2 \approx 0{,}2198$

$H_5 = 0{,}5212^2 \approx 0{,}2716$

$H_6 = 0{,}4975^2 \approx 0{,}2475$

$H_7 = 0{,}5^2 \approx 0{,}25$

$H_8 = (-0{,}5)^2 \approx 0{,}25$

=> Norm squared of J :

$$\|J\|^2 = \sum_{k=1}^{11} J_k^2$$

$$\|J\|^2 = 0{,}0029 + 0{,}0029 + 0{,}0007 + 0{,}0007 + 0{,}0010 + 0{,}0010 + 0{,}2198 + 0{,}2716 + 0{,}2475 + 0{,}25 + 0{,}25 = 1{,}2481$$

Because $H = J^T J$ is rank 1, i'm adding a damping term :

$H_{damped} = H + \lambda$ $\boxed{\text{let } \lambda = 10^{-4}}$

$$\|J\|^2 + \lambda = 1{,}2481 + 0{,}0001 = 1{,}2482$$

%% Newton update (Gauss-Newton)

derivation for a single output network update

$$\Delta w = \frac{t-y}{\|J\|^2 + \lambda} J \qquad \boxed{t - y = e}$$

$$\Delta w = \frac{e}{\|J\|^2 + \lambda}$$

=> Optimal step size

$$\alpha = \frac{e}{\|J\|^2 + \lambda} \approx \frac{-0{,}2517}{1{,}2482} \approx -0{,}2017$$

% Weight update

$\boxed{\Delta w = \alpha J}$

$\Delta w_1 = -0{,}2017 \cdot 0{,}0535 \approx -0{,}0108$

$\Delta w_9 = -0{,}2017 \cdot (-0{,}0535) \approx 0{,}0108$

$\Delta w_{11} = -0{,}2017 \cdot 0{,}0262 \approx -0{,}0053$

$\Delta w_4 = -0{,}2017 \cdot (-0{,}0262) \approx 0{,}0053$

$\Delta w_3 = -0{,}2017 \cdot (-0{,}0312) \approx \underline{\underline{0{,}0063}}$

$\Delta w_5 = -0{,}2017 \cdot 0{,}0312 \approx \underline{\underline{-0{,}0063}}$

$\Delta w_6 = -0{,}2017 \cdot 0{,}4688 \approx \underline{\underline{-0{,}0946}}$

$\Delta w_7 = -0{,}2017 \cdot 0{,}5212 \approx \underline{\underline{-0{,}1051}}$

$\Delta w_8 = -0{,}2017 \cdot 0{,}4975 \approx \underline{\underline{-0{,}1003}}$

$\Delta w_2 = -0{,}2017 \cdot 0{,}5 \approx \underline{\underline{-0{,}1009}}$

$\Delta w_{10} = -0{,}2017 \cdot (-0{,}5) \approx \underline{\underline{0{,}1009}}$

7. New updated weights $\boxed{w_{new} = w + \Delta w}$

$w_1 = -0{,}25 + (-0{,}0108) \approx -0{,}2608$

$w_2 = 0{,}33 + (-0{,}1009) \approx \underline{\underline{0{,}2291}}$

$w_3 = 0{,}14 + 0{,}0063 \approx \underline{\underline{0{,}1463}}$

$w_4 = -0{,}17 + 0{,}0053 \approx \underline{\underline{-0{,}1647}}$

$w_5 = 0{,}16 + (-0{,}0063) \approx \underline{\underline{0{,}1537}}$

$w_6 = 0{,}43 + (-0{,}0946) \approx \underline{\underline{0{,}3354}}$

$w_7 = 0{,}21 + (-0{,}1051) \approx \underline{\underline{0{,}1049}}$

$w_8 = -0{,}25 + (-0{,}1003) \approx \underline{\underline{-0{,}3503}}$

$w_9 = 0 + 0{,}0108 \approx \underline{\underline{0{,}0108}}$

$w_{10} = 0 + 0{,}1009 \approx \underline{\underline{0{,}1009}}$

$w_{11} = 0 + (-0{,}0053) \approx \underline{\underline{-0{,}0053}}$