

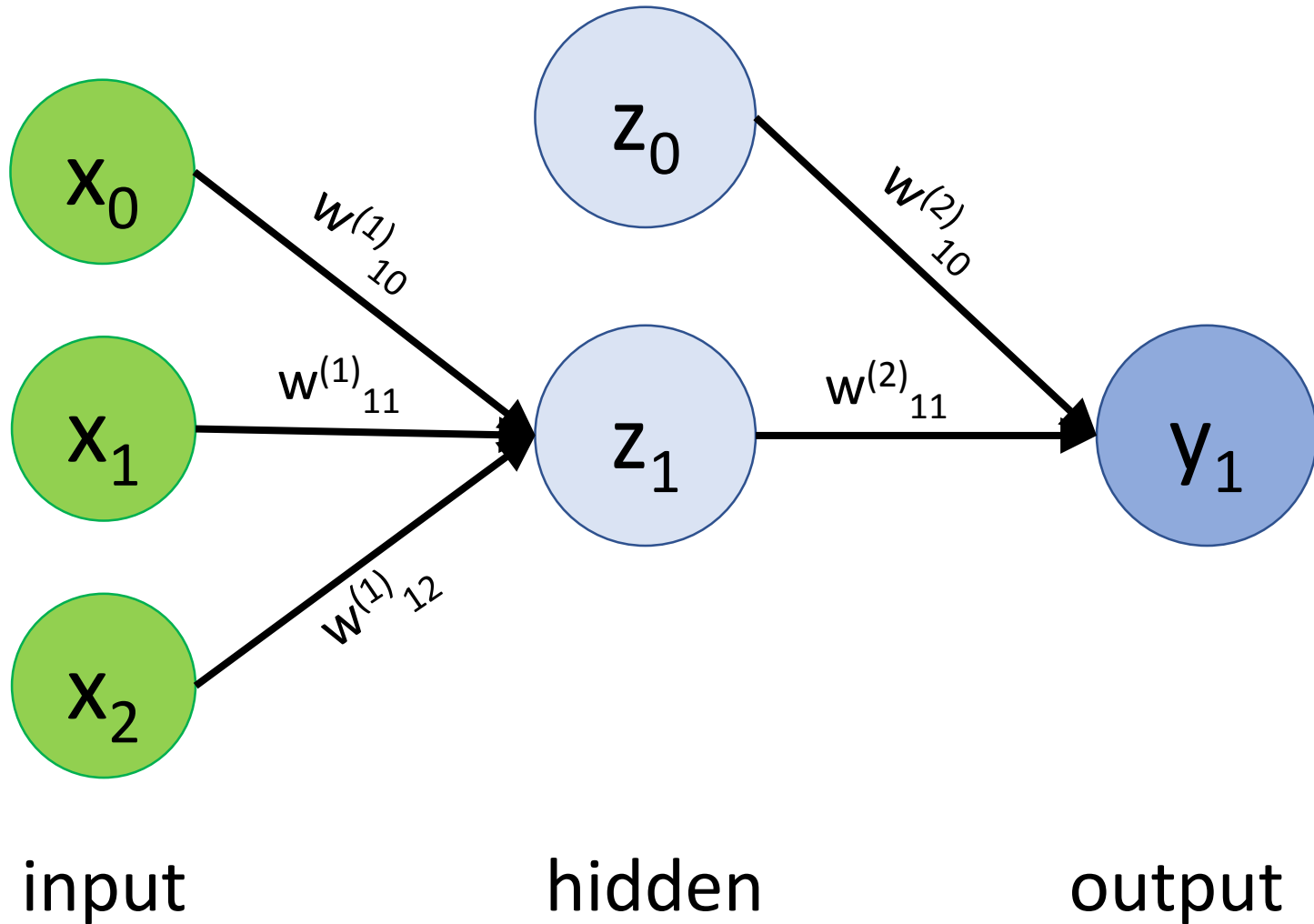
Neural Net Examples

CS 1675 / Recitation

Nov. 2, 2018

University of Pittsburgh

First architecture



Computing activations

- In all examples, $x = [x_1 \ x_2]$, plus include 1 for bias
- Assume sigmoid activation function
- Initialize all weights to 0.1
- First example: $x = [1 \ 0]$
- Second example: $x = [0 \ 1]$
- Third example: $x = [1 \ 1]$

Computing activations (answers)

- First example:

- At hidden: $z_1 = 1 / [1 + \exp(-(x_0 * w^{(1)}_{10} + x_1 * w^{(1)}_{11} + x_2 * w^{(1)}_{12}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 1 * 0.1 + 0 * 0.1))] = 0.5498$
- At output: $y_1 = 1 / [1 + \exp(-(z_0 * w^{(2)}_{10} + z_1 * w^{(2)}_{11}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 0.5498 * 0.1))] = 0.5387 \rightarrow y_{\text{pred}} = 1$

- Second example:

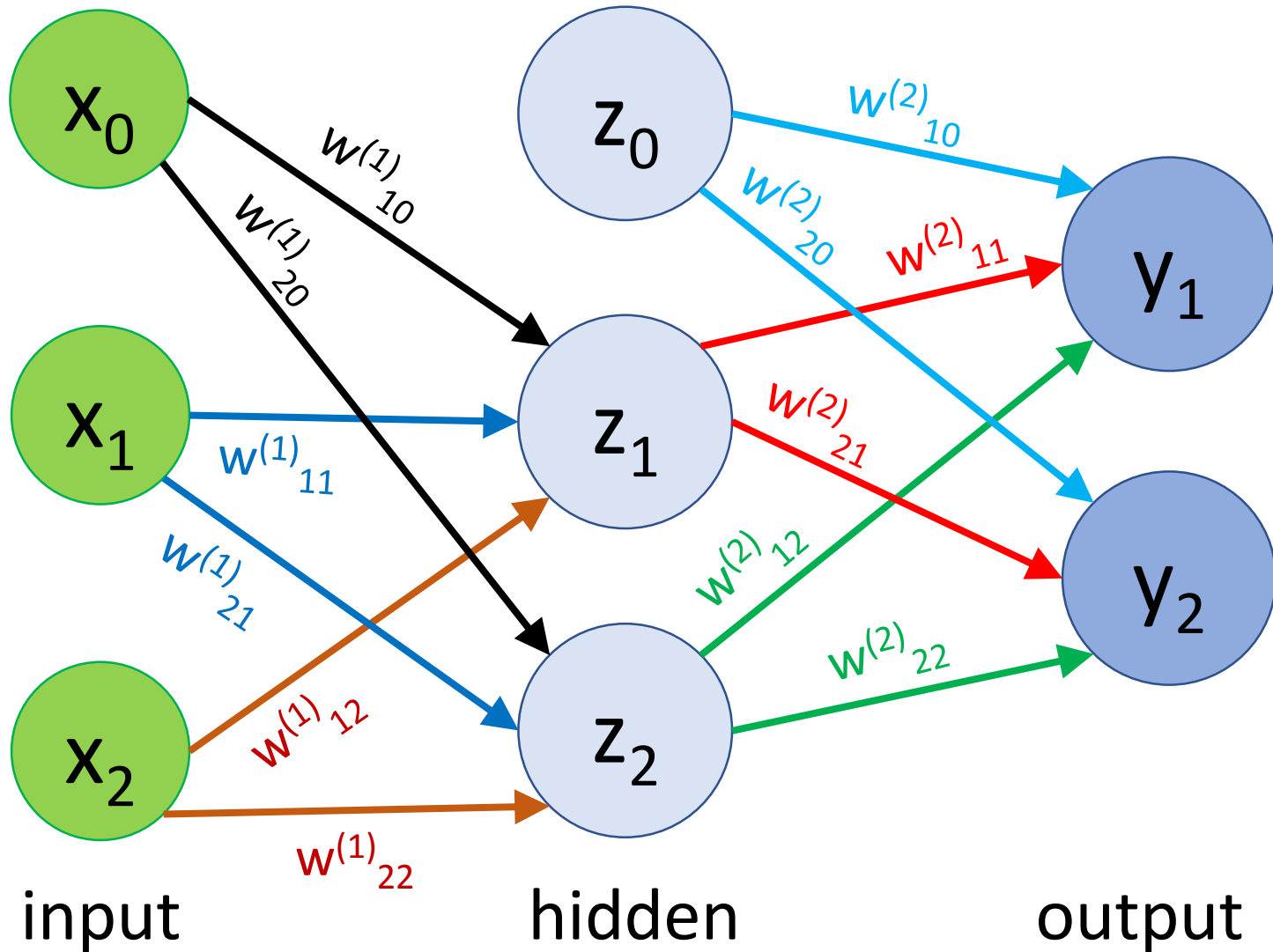
- At hidden: $z_1 = 1 / [1 + \exp(-(x_0 * w^{(1)}_{10} + x_1 * w^{(1)}_{11} + x_2 * w^{(1)}_{12}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 0 * 0.1 + 1 * 0.1))] = 0.5498$
- At output: $y_1 = 1 / [1 + \exp(-(z_0 * w^{(2)}_{10} + z_1 * w^{(2)}_{11}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 0.5498 * 0.1))] = 0.5387 \rightarrow y_{\text{pred}} = 1$

Computing activations (answers)

- Third example:

- At hidden: $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 1 * 0.1 + 1 * 0.1))] = 0.5744$
- At output: $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)}))]$
- $= 1 / [1 + \exp(-(1 * 0.1 + 0.5744 * 0.1))] = 0.5393 \rightarrow y_{\text{pred}} = 1$

Second architecture



Computing activations

- In all examples, $x = [x_1 \ x_2]$, plus include 1 for bias
- Assume sigmoid activation function
- Initialize all weights to 0.05
- First example: $x = [1 \ 0]$
- Second example: $x = [0 \ 1]$
- Third example: $x = [1 \ 1]$

Computing activations (answers)

- First example:

- At hidden:

- $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 1 * 0.05 + 0 * 0.05))] = 0.5249$

- $z_2 = 1 / [1 + \exp(-(x_0 * w_{20}^{(1)} + x_1 * w_{21}^{(1)} + x_2 * w_{22}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 1 * 0.05 + 0 * 0.05))] = 0.5249$

- At output:

- $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)} + z_2 * w_{12}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5249 * 0.05 + 0.5249 * 0.05))] =$
 0.5256

- $y_2 = 1 / [1 + \exp(-(z_0 * w_{20}^{(2)} + z_1 * w_{21}^{(2)} + z_2 * w_{22}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5249 * 0.05 + 0.5249 * 0.05))] =$
 $0.5256 \rightarrow y_{\text{pred}} = [1 \ 1]$

Computing activations (answers)

- Second example:

- At hidden:

- $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 0 * 0.05 + 1 * 0.05))] = 0.5249$

- $z_2 = 1 / [1 + \exp(-(x_0 * w_{20}^{(1)} + x_1 * w_{21}^{(1)} + x_2 * w_{22}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 0 * 0.05 + 1 * 0.05))] = 0.5249$

- At output:

- $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)} + z_2 * w_{12}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5249 * 0.05 + 0.5249 * 0.05))] =$
 0.5256

- $y_2 = 1 / [1 + \exp(-(z_0 * w_{20}^{(2)} + z_1 * w_{21}^{(2)} + z_2 * w_{22}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5249 * 0.05 + 0.5249 * 0.05))] =$
 $0.5256 \rightarrow y_{\text{pred}} = [1 \ 1]$

Computing activations (answers)

- Third example:

- At hidden:

- $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 1 * 0.05 + 1 * 0.05))] = 0.5374$

- $z_2 = 1 / [1 + \exp(-(x_0 * w_{20}^{(1)} + x_1 * w_{21}^{(1)} + x_2 * w_{22}^{(1)}))] =$
 $1 / [1 + \exp(-(1 * 0.05 + 1 * 0.05 + 1 * 0.05))] = 0.5374$

- At output:

- $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)} + z_2 * w_{12}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5374 * 0.05 + 0.5374 * 0.05))] =$
 0.5259

- $y_2 = 1 / [1 + \exp(-(z_0 * w_{20}^{(2)} + z_1 * w_{21}^{(2)} + z_2 * w_{22}^{(2)}))] =$
 $= 1 / [1 + \exp(-(1 * 0.05 + 0.5374 * 0.05 + 0.5374 * 0.05))] =$
 $0.5259 \rightarrow y_{\text{pred}} = [1 \ 1]$

Training the first network

- Perform backpropagation using stochastic gradient descent (one sample at a time)
- Weights are initially all 0.1
- Learning rate is 0.3
- Sigmoid activation function at hidden and output
- $d s(x) / dx = s(x) (1 - s(x))$
- Samples have the following labels:
 - First example: $x = [1 \ 0]$, $y = 1$
 - Second example: $x = [0 \ 1]$, $y = 0$
 - Third example: $x = [1 \ 1]$, $y = 1$
- Preview: What do you expect final weights to be?

Learning from first example

- First example: $x = [1 \ 0]$, $y = 1$
- Weights are $w_{10}^{(1)} = w_{11}^{(1)} = w_{12}^{(1)} = w_{10}^{(2)} = w_{11}^{(2)} = 0.1$
- Activations are $z_1 = 0.5498$, $y_1 = 0.5387$
- Compute errors:
 - $\delta_{y1} = y_1 * (1 - y_1) * (y_1 - y_{\text{true}}) = 0.5387 * (1 - 0.5387) * (0.5387 - 1) = -0.1146$
 - $\delta_{z1} = z_1 * (1 - z_1) * (w_{11}^{(2)} * \delta_{y1}) = 0.5498 * (1 - 0.5498) * [0.1 * -0.1146]$
 $= -0.0028$
- Update weights:
 - $w_{10}^{(2)} = w_{10}^{(2)} - 0.3 * \delta_{y1} * z_0 = 0.1 + 0.3 * 0.1146 * 1 = 0.1343$
 - $w_{11}^{(2)} = w_{11}^{(2)} - 0.3 * \delta_{y1} * z_1 = 0.1 + 0.3 * 0.1146 * 0.5498 = 0.1189$
 - $w_{10}^{(1)} = w_{10}^{(1)} - 0.3 * \delta_{z1} * x_0 = 0.1 + 0.3 * 0.0028 * 1 = 0.1008$
 - $w_{11}^{(1)} = w_{11}^{(1)} - 0.3 * \delta_{z1} * x_1 = 0.1 + 0.3 * 0.0028 * 1 = 0.1008$
 - $w_{12}^{(1)} = w_{12}^{(1)} - 0.3 * \delta_{z1} * x_2 = 0.1 + 0.3 * 0.0028 * 0 = 0.1$

Learning from second example

- Second example: $x = [0 \ 1]$, $y = 0$
- Weights are $w_{10}^{(1)} = w_{11}^{(1)} = 0.1008$, $w_{12}^{(1)} = 0.1$, $w_{10}^{(2)} = 0.1343$, $w_{11}^{(2)} = 0.1189$
- Activations are (recompute with new weights):
 - $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))] = 1 / [1 + \exp(-(1 * 0.1008 + 0 * 0.1008 + 1 * 0.1))] = 0.55$
 - $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)}))] = 1 / [1 + \exp(-(1 * 0.1343 + 0.55 * 0.1189))] = 0.5498$
- Compute errors:
 - $\delta_{y1} = y_1 * (1 - y_1) * (y_1 - y_{\text{true}}) = 0.5498 * (1 - 0.5498) * (0.5498 - 0) = 0.1361$
 - $\delta_{z1} = z_1 * (1 - z_1) * (w_{11}^{(2)} * \delta_{y1}) = 0.55 * (1 - 0.55) * [0.1189 * 0.1361] = 0.004$
- Update weights:
 - $w_{10}^{(2)} = w_{10}^{(2)} - 0.3 * \delta_{y1} * z_0 = 0.1343 - 0.3 * 0.1361 * 1 = 0.0935$
 - $w_{11}^{(2)} = w_{11}^{(2)} - 0.3 * \delta_{y1} * z_1 = 0.1189 - 0.3 * 0.1361 * 0.55 = 0.0964$
 - $w_{10}^{(1)} = w_{10}^{(1)} - 0.3 * \delta_{z1} * x_0 = 0.1008 - 0.3 * 0.004 * 1 = 0.0996$
 - $w_{11}^{(1)} = w_{11}^{(1)} - 0.3 * \delta_{z1} * x_1 = 0.1008 - 0.3 * 0.004 * 0 = 0.1008$
 - $w_{12}^{(1)} = w_{12}^{(1)} - 0.3 * \delta_{z1} * x_2 = 0.1 - 0.3 * 0.004 * 1 = 0.0988$

Learning from third example

- Third example: $x = [1 \ 1]$, $y = 1$
- Weights are $w_{10}^{(1)} = 0.0996$, $w_{11}^{(1)} = 0.1008$, $w_{12}^{(1)} = 0.0988$,
 $w_{10}^{(2)} = 0.0935$, $w_{11}^{(2)} = 0.0964$
- Activations are (recompute with new weights):
 - $z_1 = 1 / [1 + \exp(-(x_0 * w_{10}^{(1)} + x_1 * w_{11}^{(1)} + x_2 * w_{12}^{(1)}))] = 1 / [1 + \exp(-(1 * 0.0996 + 1 * 0.1008 + 1 * 0.0988))] = 0.5742$
 - $y_1 = 1 / [1 + \exp(-(z_0 * w_{10}^{(2)} + z_1 * w_{11}^{(2)}))] = 1 / [1 + \exp(-(1 * 0.0935 + 0.5735 * 0.0964))] = 0.5371$
- Compute errors:
 - $\delta_{y1} = y_1 * (1 - y_1) * (y_1 - y_{\text{true}}) = 0.5371 * (1 - 0.5371) * (0.5371 - 1) = -0.1151$
 - $\delta_{z1} = z_1 * (1 - z_1) * (w_{11}^{(2)} * \delta_{y1}) = 0.5742 * (1 - 0.5742) * [0.0964 * -0.1151] = -0.0027$
- Update weights:
 - $w_{10}^{(2)} = w_{10}^{(2)} - 0.3 * \delta_{y1} * z_0 = 0.0935 + 0.3 * 0.1151 * 1 = 0.1280$
 - $w_{11}^{(2)} = w_{11}^{(2)} - 0.3 * \delta_{y1} * z_1 = 0.0964 + 0.3 * 0.1151 * 0.5735 = 0.1162$
 - $w_{10}^{(1)} = w_{10}^{(1)} - 0.3 * \delta_{z1} * x_0 = 0.0996 + 0.3 * 0.0027 * 1 = 0.1004$
 - $w_{11}^{(1)} = w_{11}^{(1)} - 0.3 * \delta_{z1} * x_1 = 0.1008 + 0.3 * 0.0027 * 1 = 0.1016$
 - $w_{12}^{(1)} = w_{12}^{(1)} - 0.3 * \delta_{z1} * x_2 = 0.0988 + 0.3 * 0.0027 * 1 = 0.0996$

Recap

- Do the $w^{(1)}$ weights we obtained make sense?