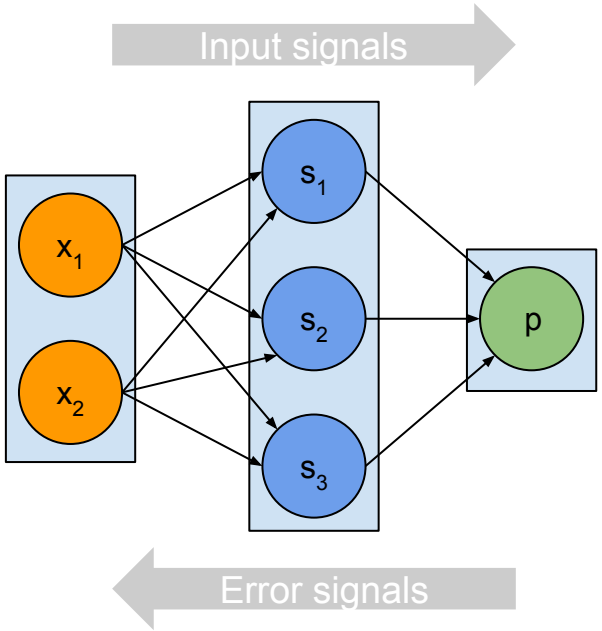
# Training a neural network: Backward propagation and gradient descent

Valerio Velardo

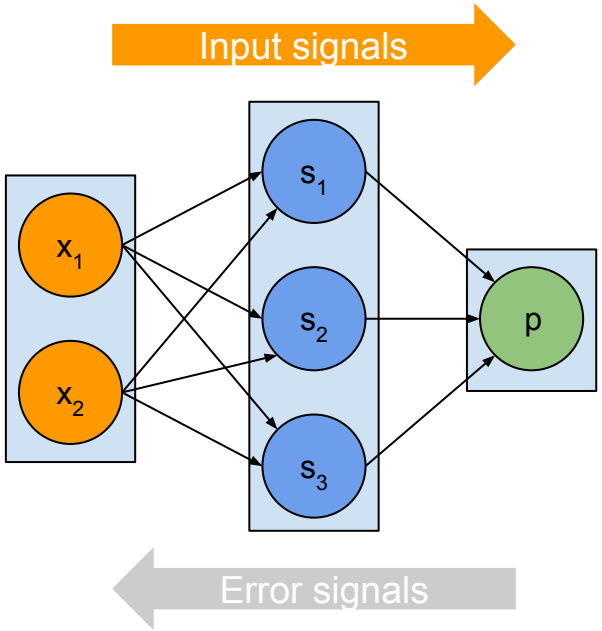
# Training a neural network

- Tweak weights of the connections
- Feed training data (input + target) to the network
- Iterative adjustments

# Training a neural network

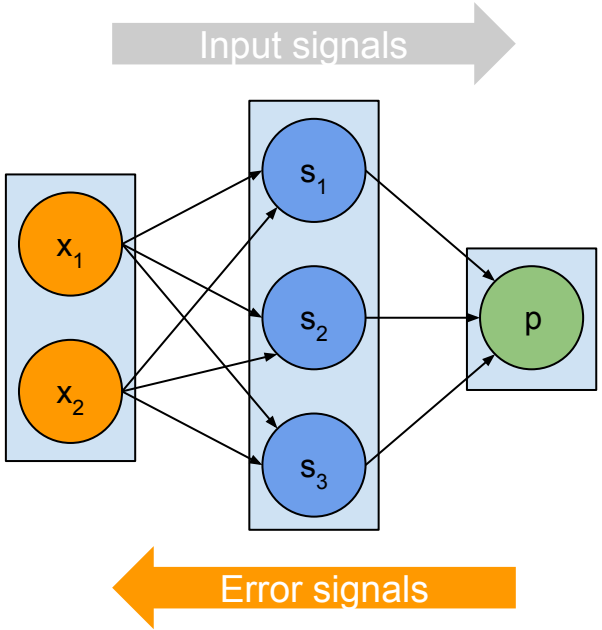


# Training a neural network



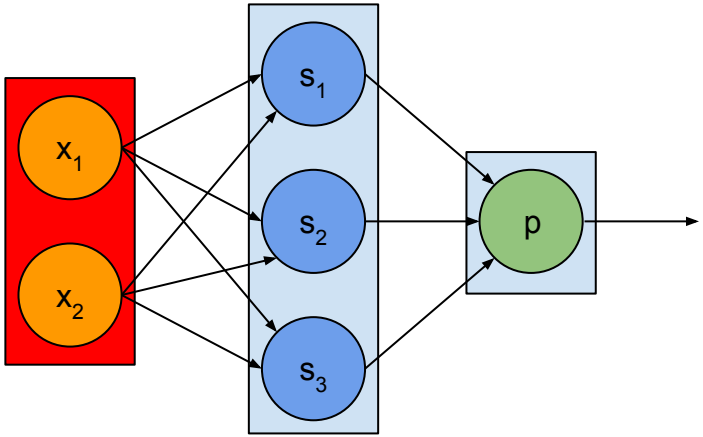

1. Get prediction
2. Calculate error
3. Calculate gradient of error function over the weights
4. Update parameters

# Training a neural network

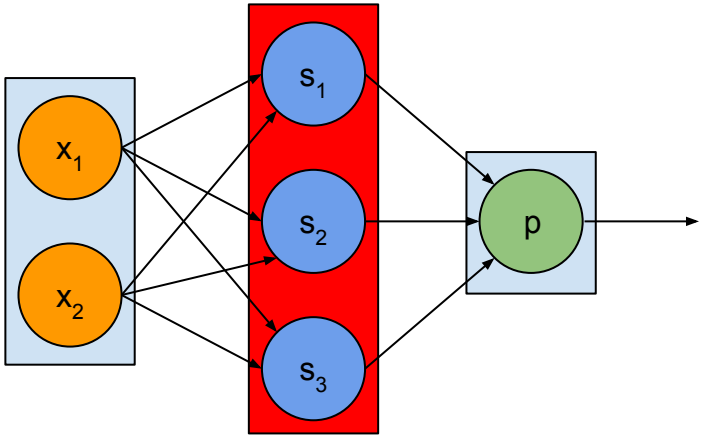



1. Get prediction
2. Calculate error
3. Calculate gradient of error function over the weights
4. Update parameters

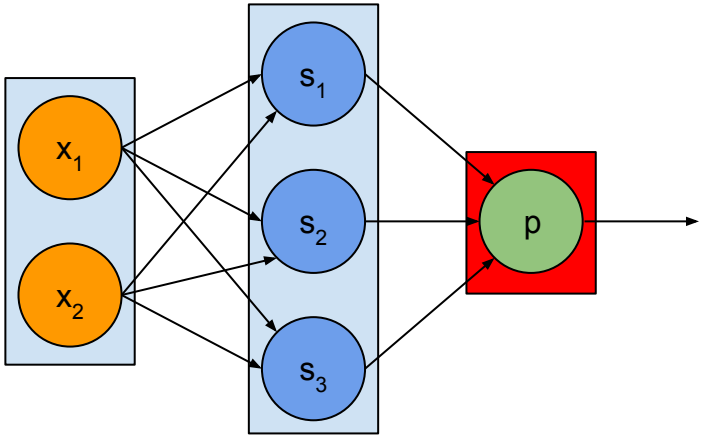
# Get prediction



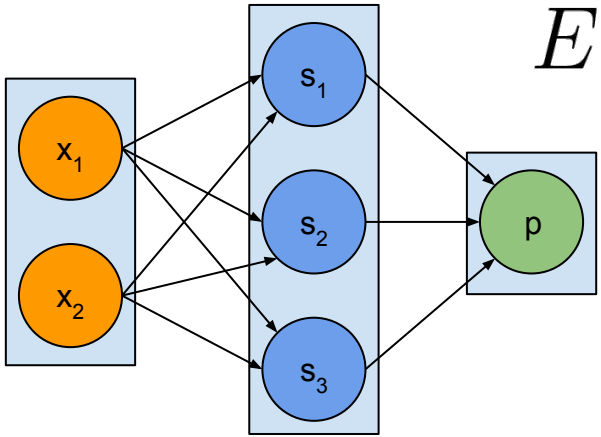
# Get prediction



# Get prediction

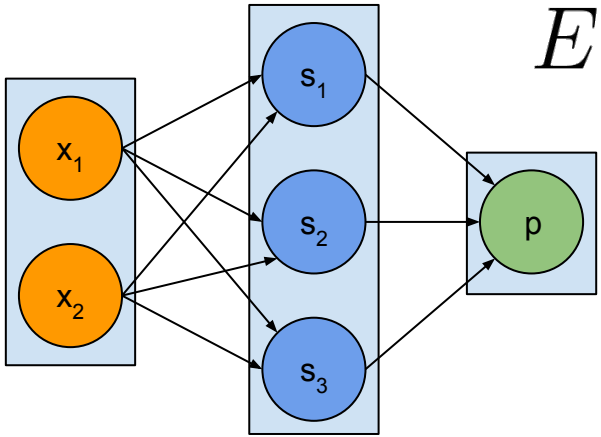


# Calculate error



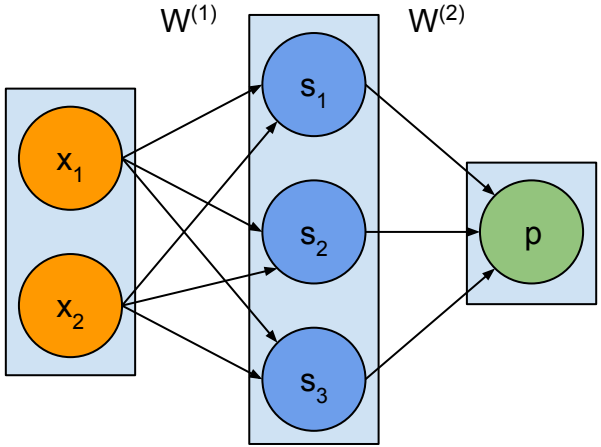

$$E = E(\mathbf{p}, \mathbf{y})$$

# Calculate error



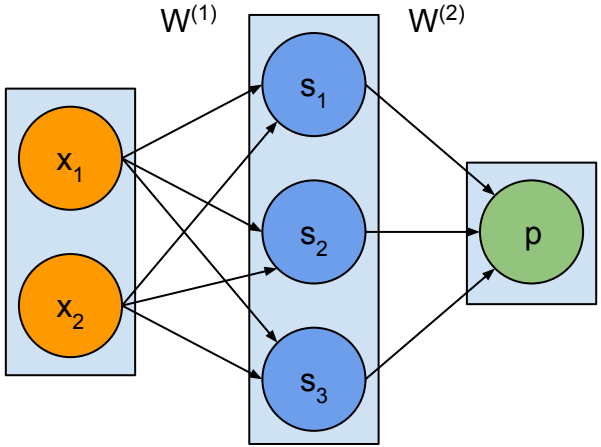

$$E = E(\mathbf{p}, \mathbf{y}) = \frac{1}{2}(\mathbf{p} - \mathbf{y})^2$$

# Calculate gradient of error function



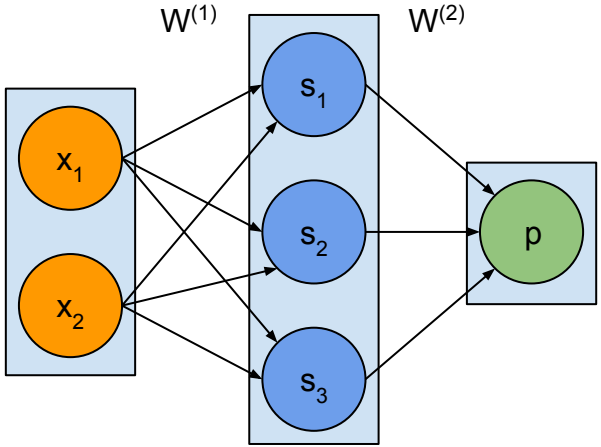

$$\frac{\partial E}{\partial W^{(n)}}$$

# Calculate gradient of error function




$$F = F(\mathbf{x}, W)$$

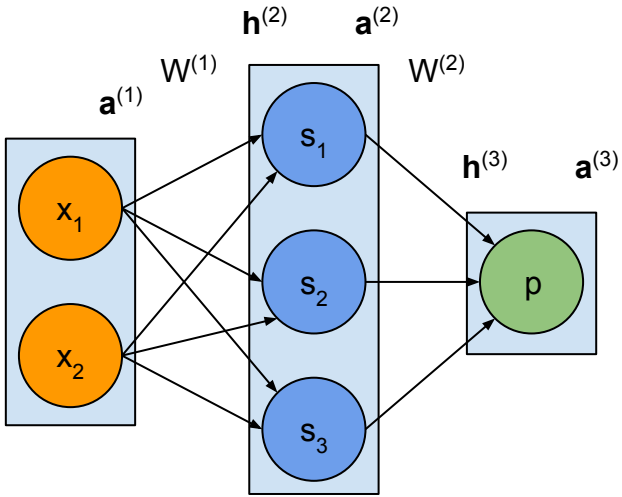
# Calculate gradient of error function



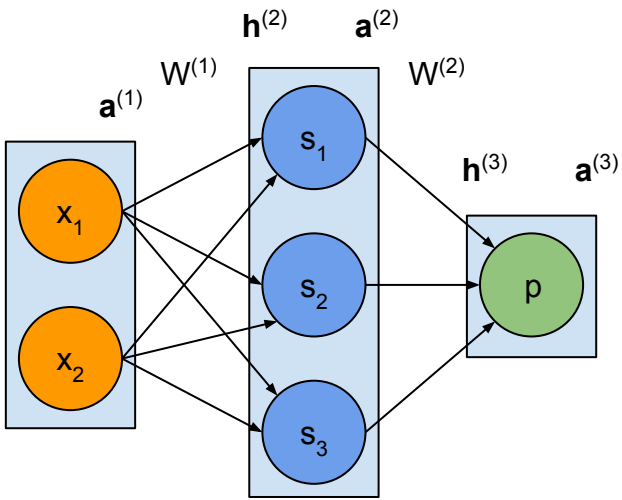

$$F = F(\mathbf{x}, W)$$

$$E = E(\mathbf{p}, \mathbf{y}) = E(F(\mathbf{x}, W), \mathbf{y})$$

# Calculate gradient of error function

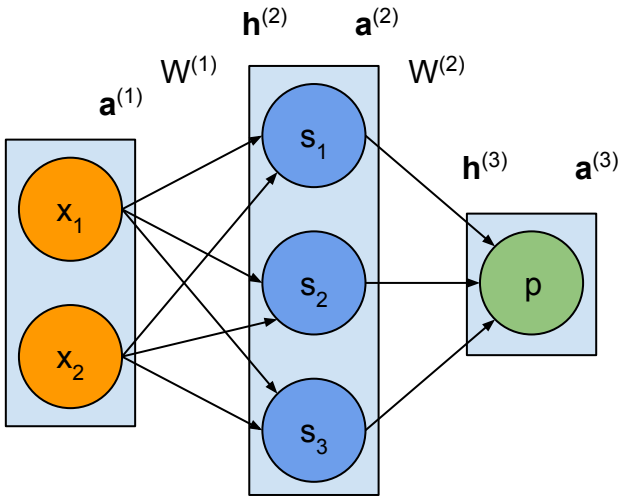


# Calculate gradient of error function



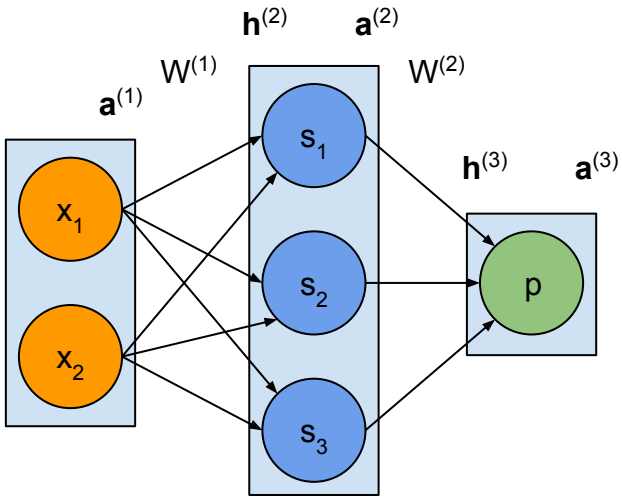

$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \boxed{\frac{\partial E}{\partial a^{(3)}}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

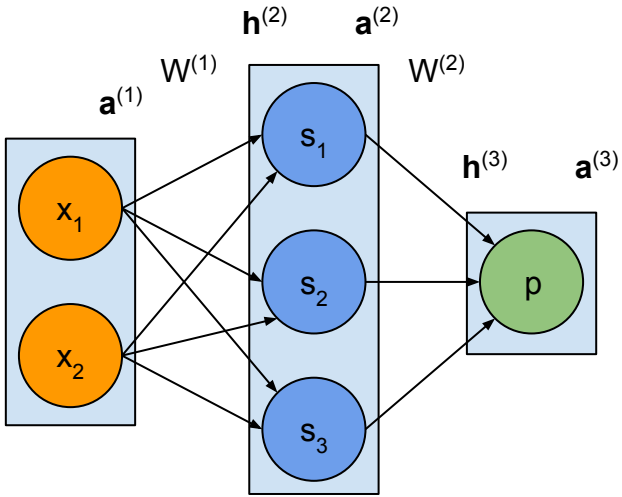
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$E = \frac{1}{2}(a^{(3)} - y)^2$$

# Calculate gradient of error function

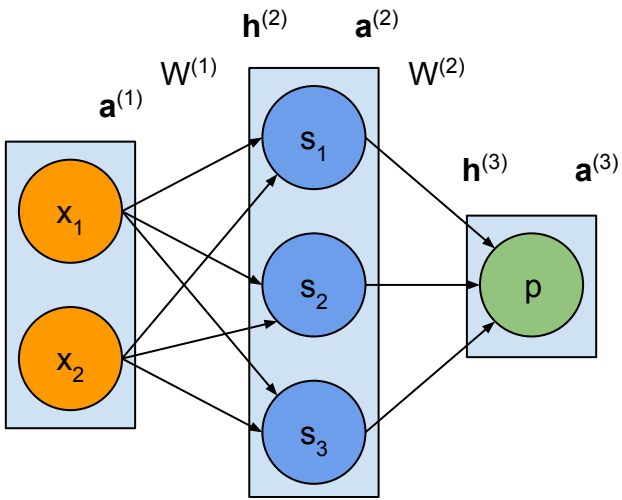



$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$E = \frac{1}{2}(a^{(3)} - y)^2$$

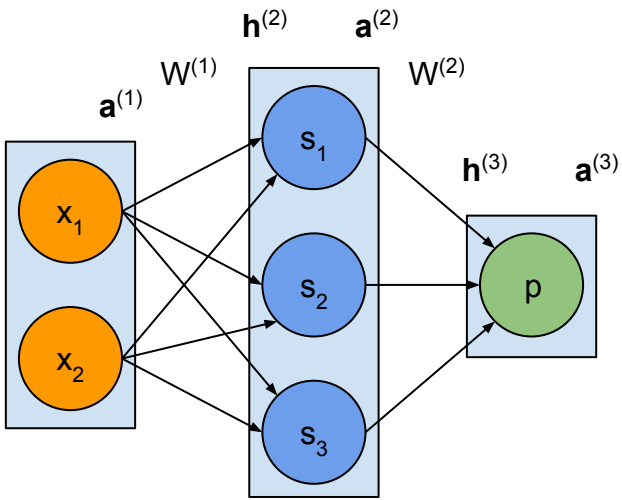
$$\frac{\partial E}{\partial a^{(3)}} = 2 \cdot \frac{1}{2}(a^{(3)} - y) \cdot 1 = a^{(3)} - y$$

# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

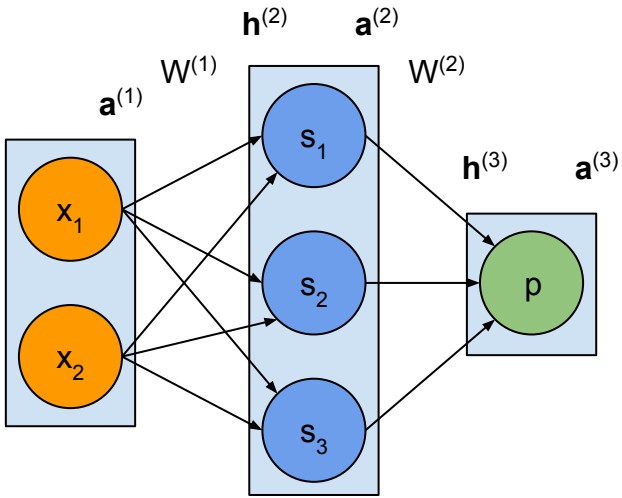
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$a^{(3)} = \frac{1}{1 + e^{-h^{(3)}}}$$

# Calculate gradient of error function

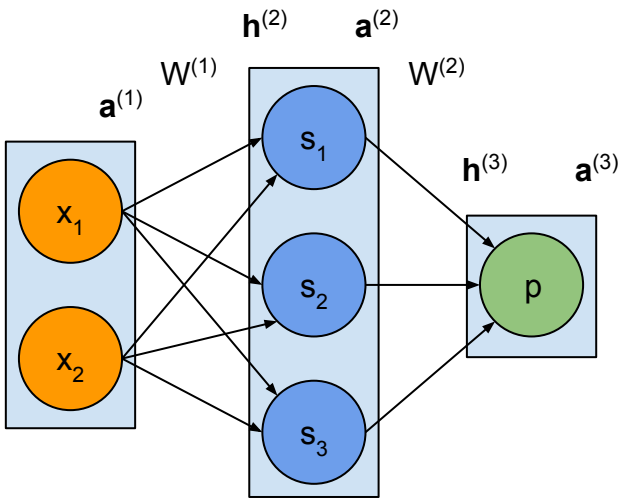



$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$a^{(3)} = \frac{1}{1 + e^{-h^{(3)}}}$$

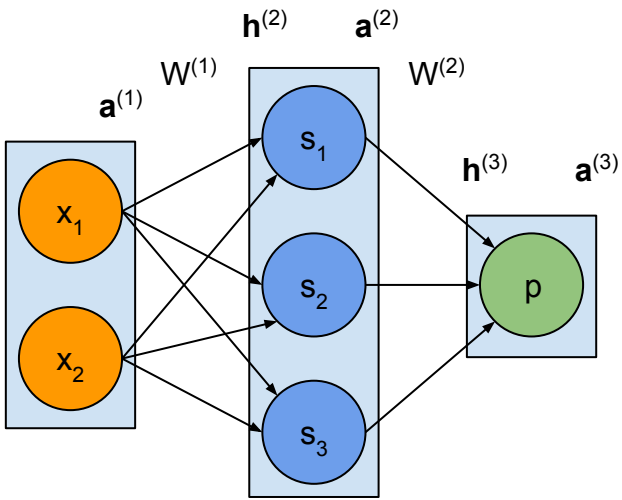
$$\frac{\partial a^{(3)}}{\partial h^{(3)}} = \sigma'(h^{(3)}) = \sigma(h^{(3)})(1-\sigma(h^{(3)}))$$

# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial W^{(2)}}}$$

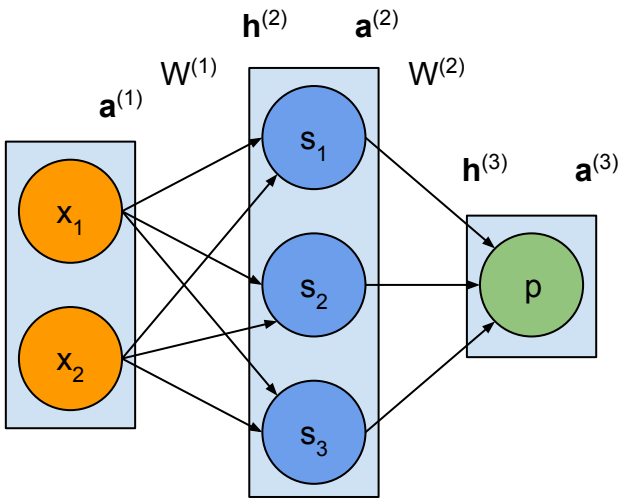
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$h^{(3)} = a^{(2)} W^{(2)}$$

# Calculate gradient of error function

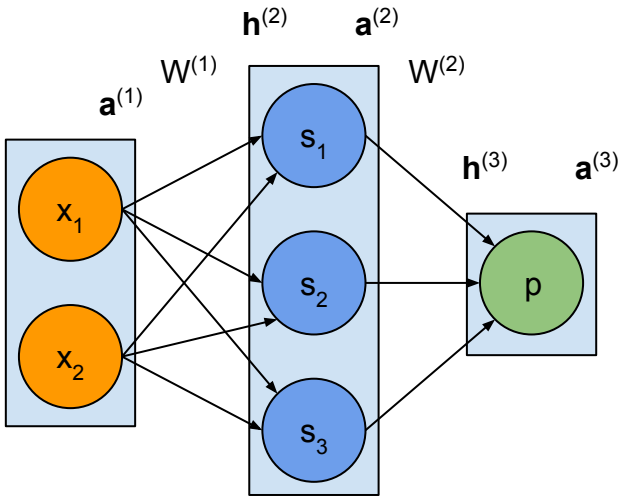



$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial W^{(2)}}}$$

$$h^{(3)} = a^{(2)} W^{(2)}$$

$$\boxed{\frac{\partial h^{(3)}}{\partial W^{(2)}} = a^{(2)}}$$

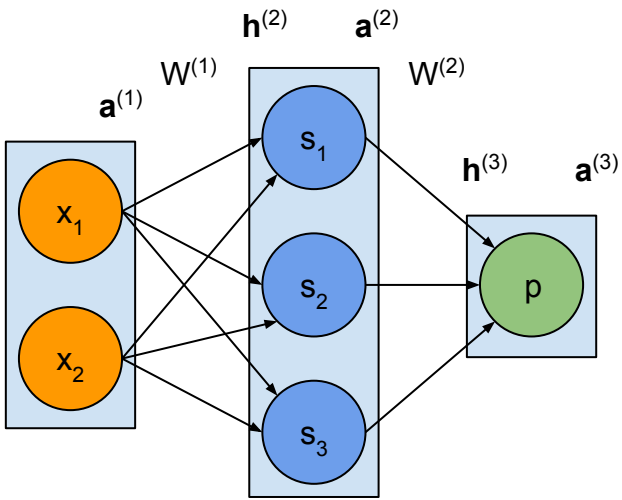
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$\frac{\partial E}{\partial W^{(2)}} =$$

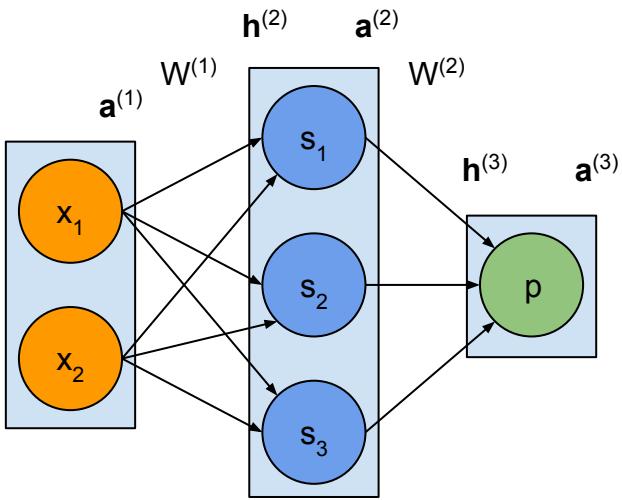
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \boxed{\frac{\partial E}{\partial a^{(3)}}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$\frac{\partial E}{\partial W^{(2)}} = \boxed{(a^{(3)} - y)}$$

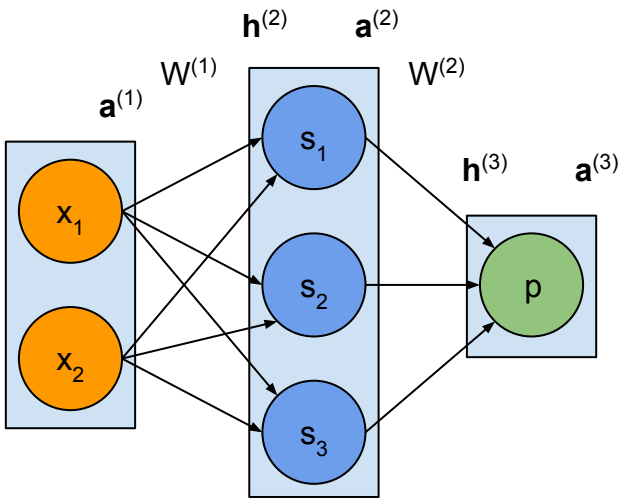
# Calculate gradient of error function




$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial W^{(2)}}$$

$$\frac{\partial E}{\partial W^{(2)}} = (a^{(3)} - y)\sigma'(h^{(3)})$$

# Calculate gradient of error function

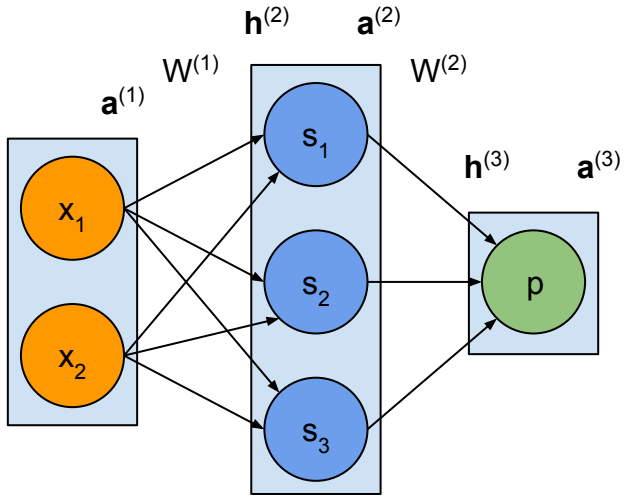



$$\frac{\partial E}{\partial W^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial W^{(2)}}}$$

$$\frac{\partial E}{\partial W^{(2)}} = (a^{(3)} - y)\sigma'(h^{(3)})\boxed{a^{(2)}}$$

# Calculate gradient of error function

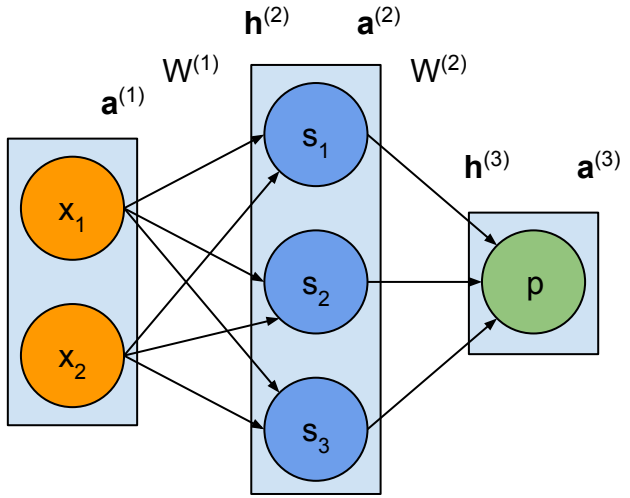
$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$\frac{\partial E}{\partial a^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}}$$

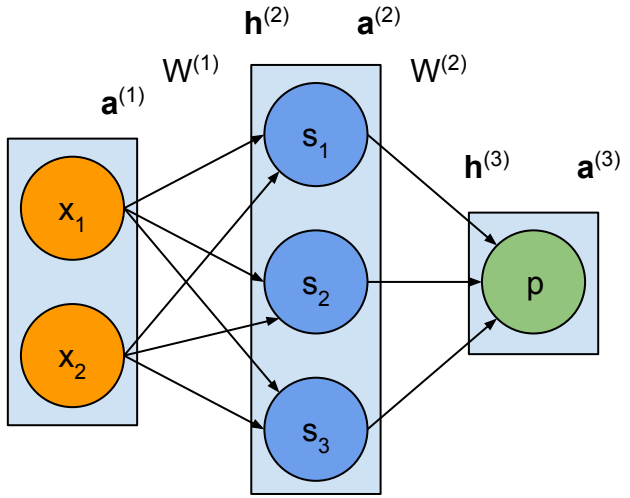


# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

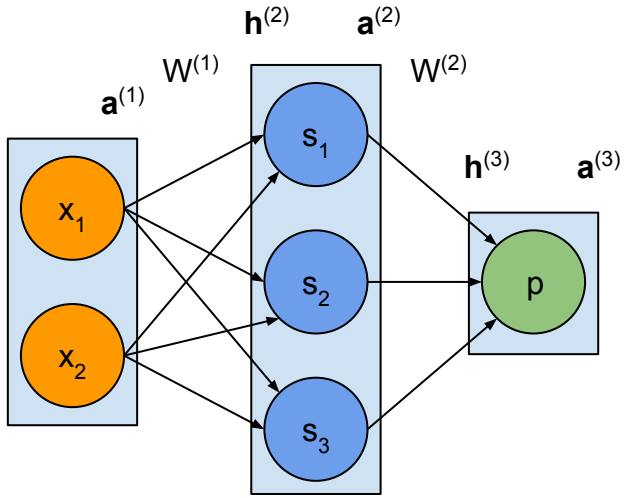
$$\frac{\partial E}{\partial a^{(2)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}}$$

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$



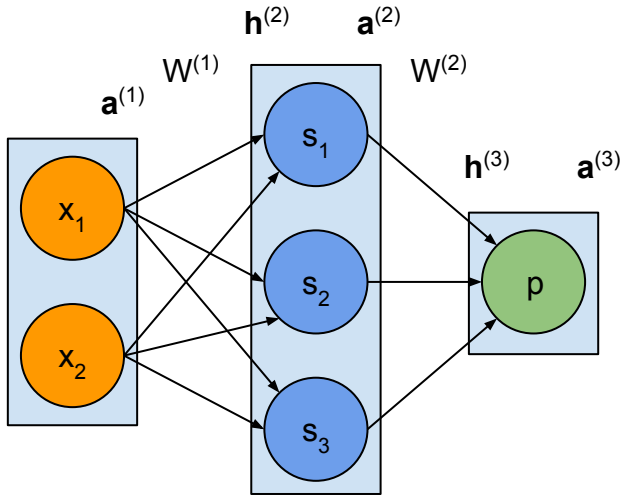
# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}}$$



# Calculate gradient of error function

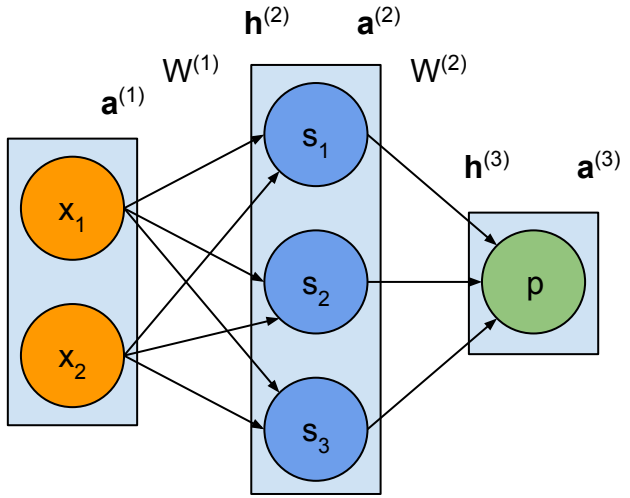
$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial a^{(2)}}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$h^{(3)} = a^{(2)} W^{(2)}$$

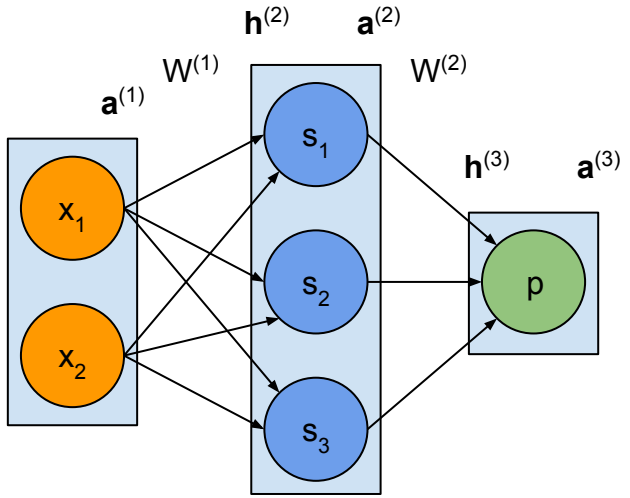


# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial a^{(2)}}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

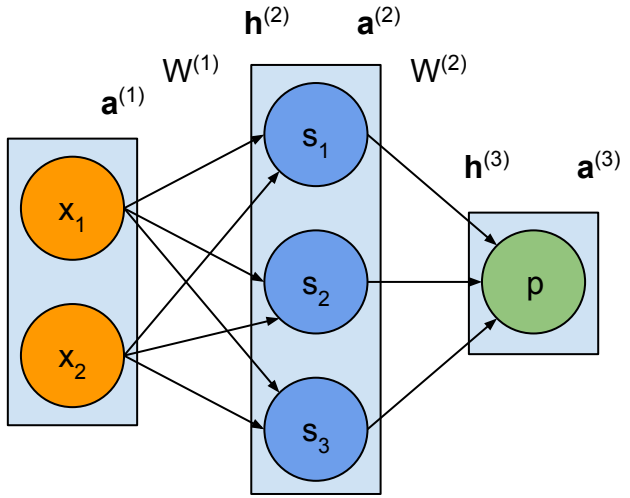
$$h^{(3)} = a^{(2)} W^{(2)}$$

$$\boxed{\frac{\partial h^{(3)}}{\partial a^{(2)}} = W^{(2)}}$$



# Calculate gradient of error function

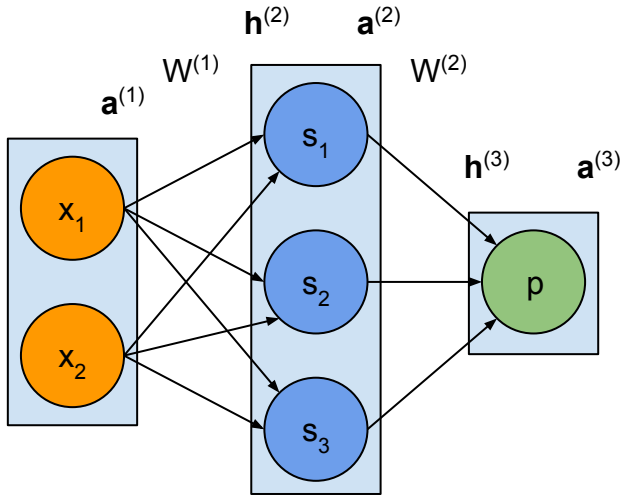
$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \boxed{\frac{\partial a^{(2)}}{\partial h^{(2)}}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$a^{(2)} = \frac{1}{1 + e^{-h^{(2)}}}$$

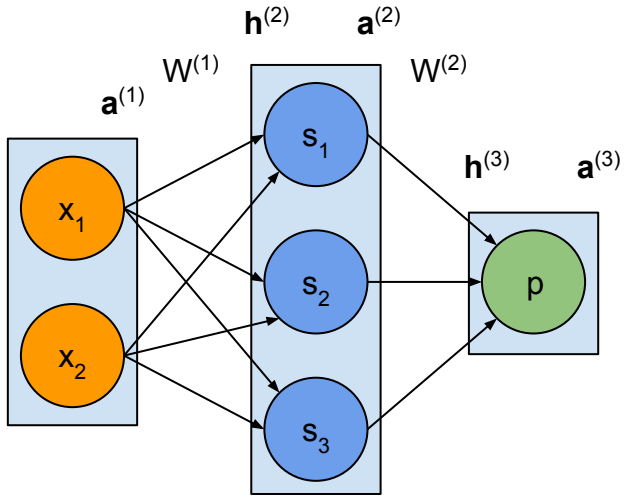


# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \boxed{\frac{\partial a^{(2)}}{\partial h^{(2)}}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

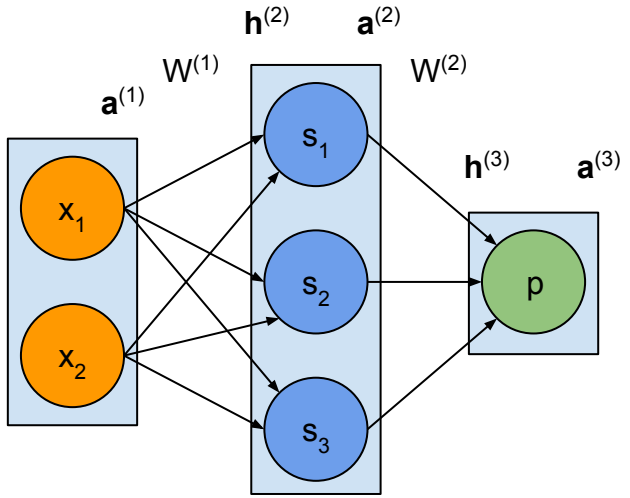
$$a^{(2)} = \frac{1}{1 + e^{-h^{(2)}}}$$

$$\boxed{\frac{\partial a^{(2)}}{\partial h^{(2)}} = \sigma'(h^{(2)}) = \sigma(h^{(2)})(1 - \sigma(h^{(2)}))}$$



# Calculate gradient of error function

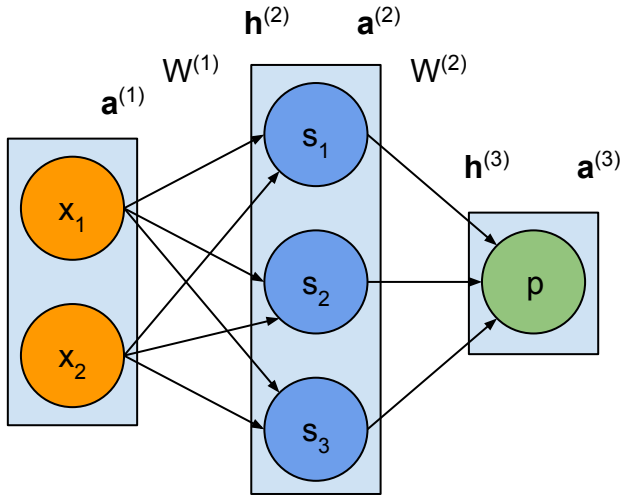
$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \boxed{\frac{\partial h^{(2)}}{\partial W^{(1)}}}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$h^{(2)} = xW^{(1)}$$

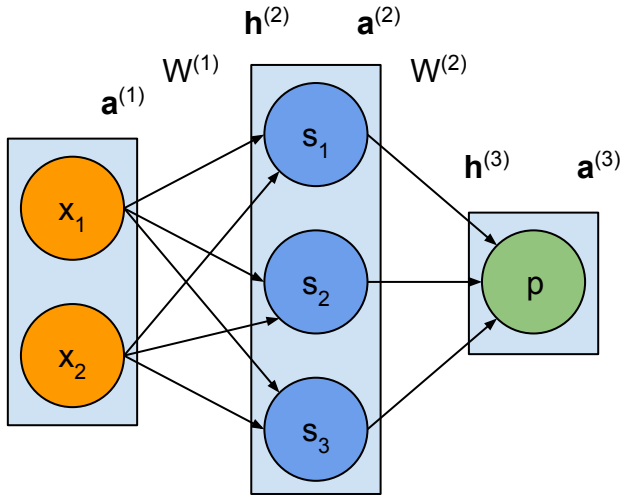


# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$h^{(2)} = xW^{(1)}$$

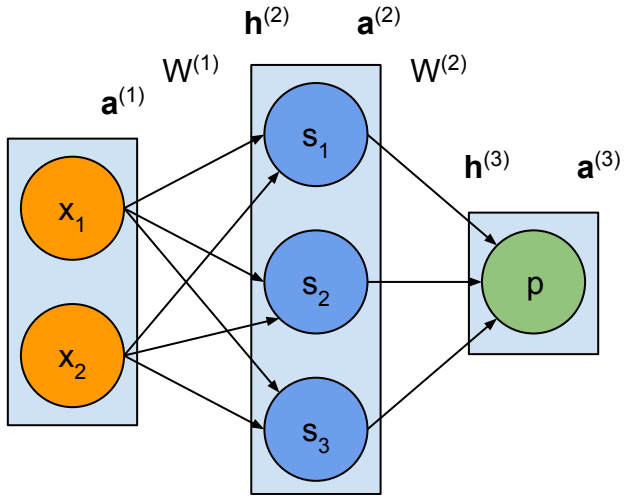
$$\frac{\partial h^{(2)}}{\partial W^{(1)}} = x$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

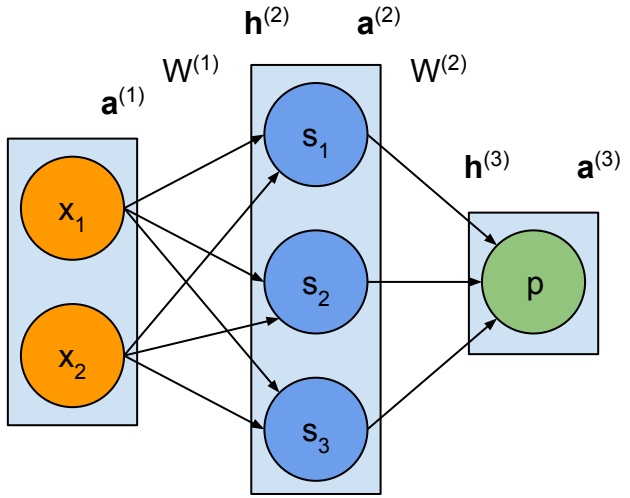
$$\frac{\partial E}{\partial W^{(1)}} =$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

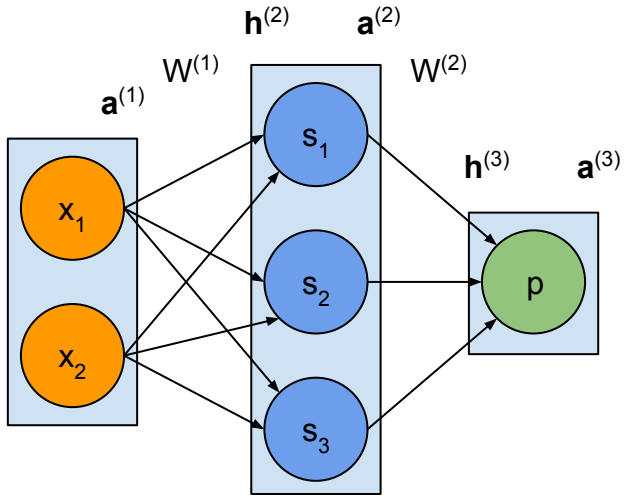
$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

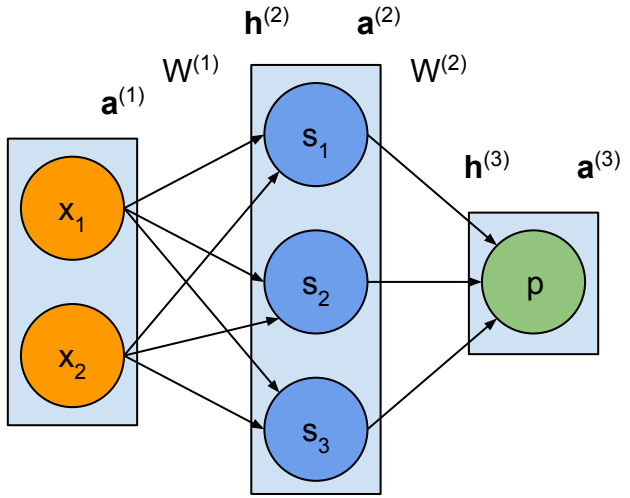
$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)\sigma'(h^{(3)})$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \boxed{\frac{\partial h^{(3)}}{\partial a^{(2)}}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

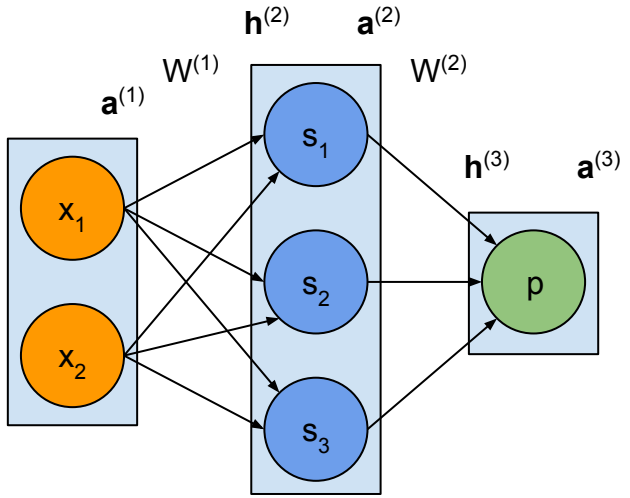
$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)\sigma'(h^{(3)}) \boxed{W^{(2)}}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \boxed{\frac{\partial a^{(2)}}{\partial h^{(2)}}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

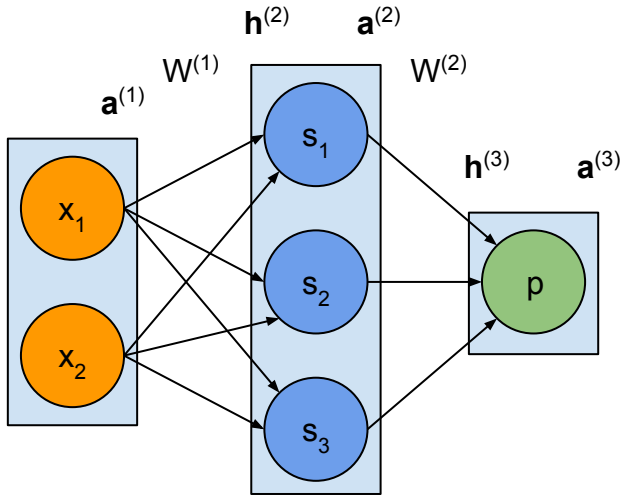
$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)\sigma'(h^{(3)})W^{(2)}\boxed{\sigma'(h^{(2)})}$$



# Calculate gradient of error function

$$\frac{\partial E}{\partial W^{(1)}} = \frac{\partial E}{\partial a^{(3)}} \frac{\partial a^{(3)}}{\partial h^{(3)}} \frac{\partial h^{(3)}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial h^{(2)}} \frac{\partial h^{(2)}}{\partial W^{(1)}}$$

$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)\sigma'(h^{(3)})W^{(2)}\sigma'(h^{(2)})x$$

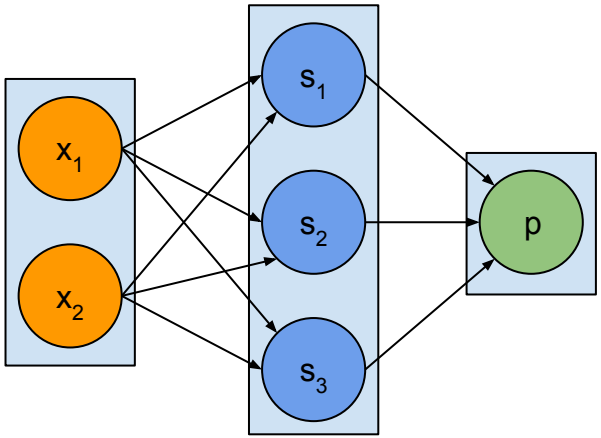


# Calculate gradient of error function

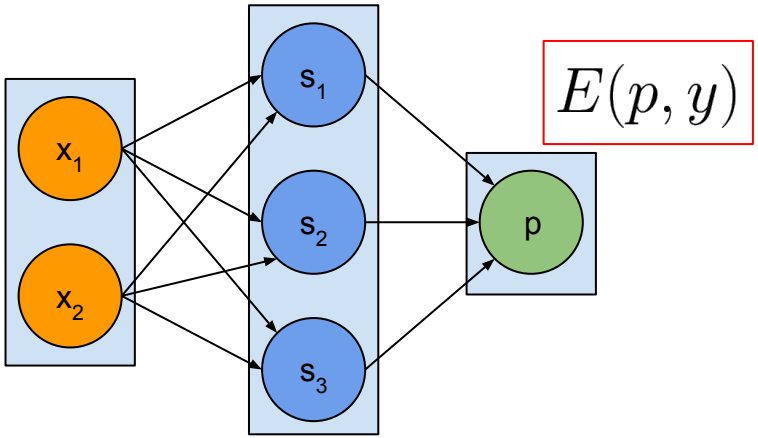
$$\frac{\partial E}{\partial W^{(2)}} = (a^{(3)} - y)\sigma'(h^{(3)})a^{(2)}$$

$$\frac{\partial E}{\partial W^{(1)}} = (a^{(3)} - y)\sigma'(h^{(3)})W^{(2)}\sigma'(h^{(2)})x$$

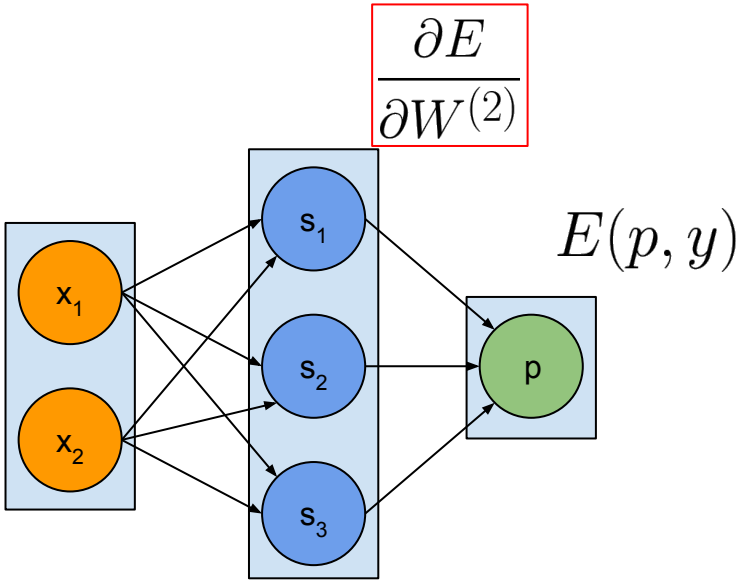
# Backpropagation



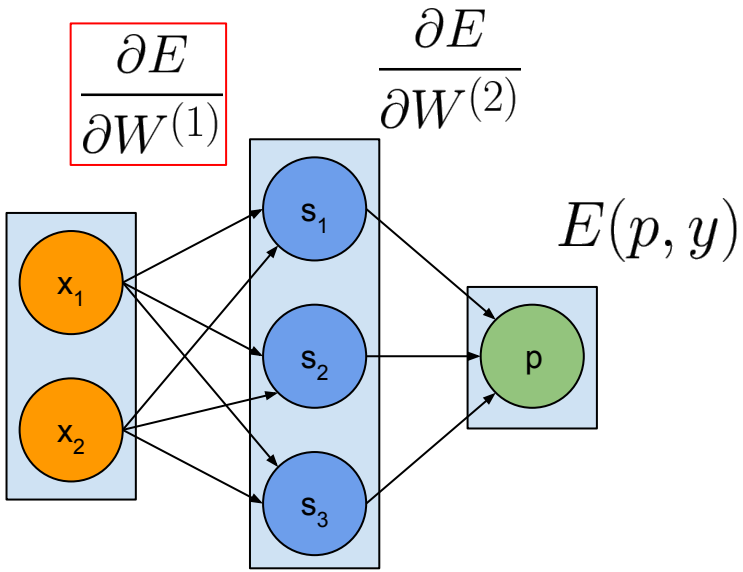
# Backpropagation



# Backpropagation



# Backpropagation

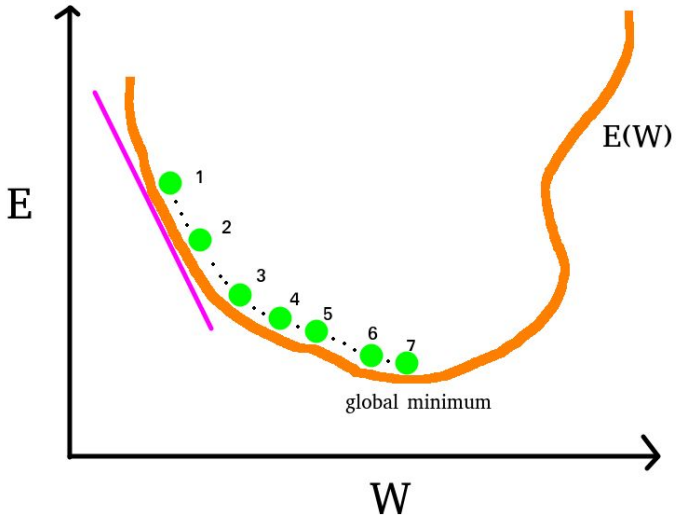


# Training steps

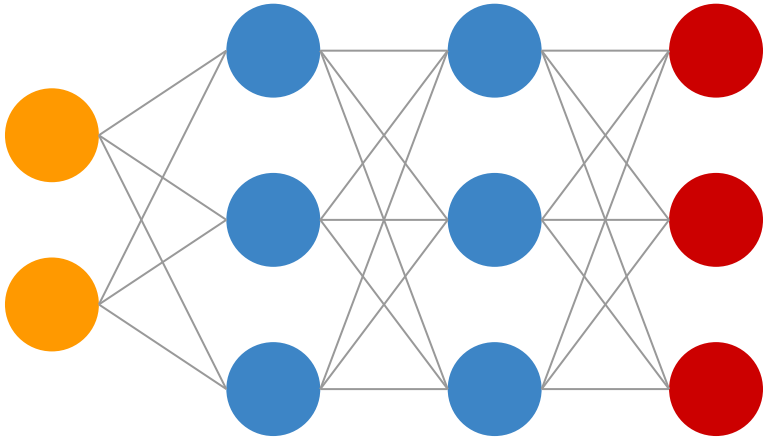
1. Get prediction
2. Calculate error
3. Calculate gradient of error function over the weights
4. Update parameters

# Gradient descent

- Take a step in opposite direction to gradient
- Step = Learning rate



# What’s up next?



python™