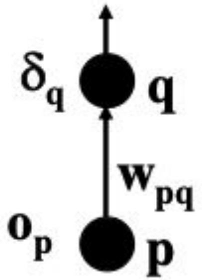
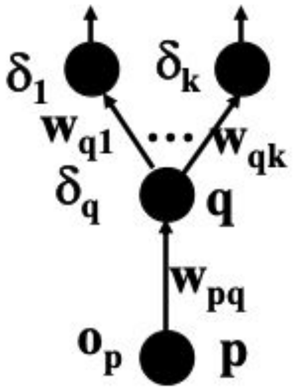
Backpropagation Rule - Summary

$w_{pq}(t)$ - weight from node p to node q at time t

$$w_{pq}(t+1) = w_{pq}(t) + \Delta w_{pq}$$

$$\Delta w_{pq} = \eta \cdot \delta_q \cdot o_p$$ - weight change





- The weight change is proportional to the output activation of neuron p and the error δ of neuron q
- δ is calculated in 2 different ways:
  - q is an output neuron $\delta_q = (d_q - o_q) \cdot f'(net_q)$
  - q is a hidden neuron $\delta_q = f'(net_q) \sum_i w_{qi} \delta_i$ (i is over the nodes in the layer above q)

Derivative of the activation function used in neuron q with respect to the input of q (netq)

Exercise 2

a) Given is a 2-layer feed-forward neural network with 10 input units, 5 hidden units and 3 output units. How many weights does it contain? Don't forget to include the bias weights.

Number of weights = ?

Number of biases = ?

Total = weights + biases = ?

Exercise 2

a) Given is a 2-layer feed-forward neural network with 10 input units, 5 hidden units and 3 output units. How many weights does it contain? Don't forget to include the bias weights.

Input to hidden

Hidden to output

Number of weights = (10 * 5) + (5 * 3) = 65

Number of biases = ?

Total = weights + biases = ?

Exercise 2

a) Given is a 2-layer feed-forward neural network with 10 input units, 5 hidden units and 3 output units. How many weights does it contain? Don't forget to include the bias weights.

Input to hidden

Hidden to output

Number of weights = $\underbrace{(10 * 5)}_{\text{Input to hidden}} + \underbrace{(5 * 3)}_{\text{Hidden to output}} = 65$

Number of biases = $5 + 3 = 8$

Total = weights + biases = ?

Exercise 2

a) Given is a 2-layer feed-forward neural network with 10 input units, 5 hidden units and 3 output units. How many weights does it contain? Don't forget to include the bias weights.

Number of weights = $\overset{\text{Input to hidden}}{\underline{(10 * 5)}} + \overset{\text{Hidden to output}}{\underline{(5 * 3)}} = 65$

Number of biases = $5 + 3 = 8$

Total = weights + biases = $65 + 8 = 73$

Exercise 2

b) Is the backpropagation algorithm guaranteed to achieve 100% correct classification for any linearly-separable set of training examples, given a sufficiently small learning rate? Explain briefly.

Exercise 2

b) Is the backpropagation algorithm guaranteed to achieve 100% correct classification for any linearly-separable set of training examples, given a sufficiently small learning rate? Explain briefly.

Answer: **No, it will iterate until a local minimum of the error is reached.**

Exercise 3

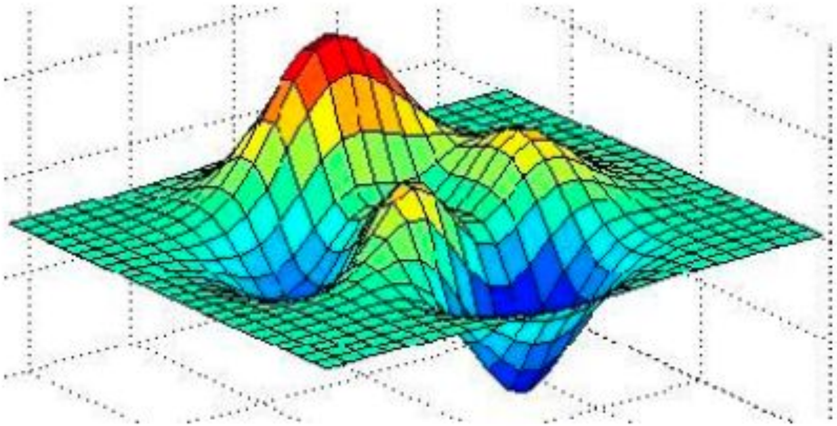
Question: Cybenko’s theorem (slide 39) states that any continuous function can be approximated by a backpropagation network with 1 hidden layer. Why do we use networks with more than 1 hidden layer?

Exercise 3

Question: Cybenko’s theorem (slide 39) states that any continuous function can be approximated by a backpropagation network with 1 hidden layer. Why do we use networks with more than 1 hidden layer?

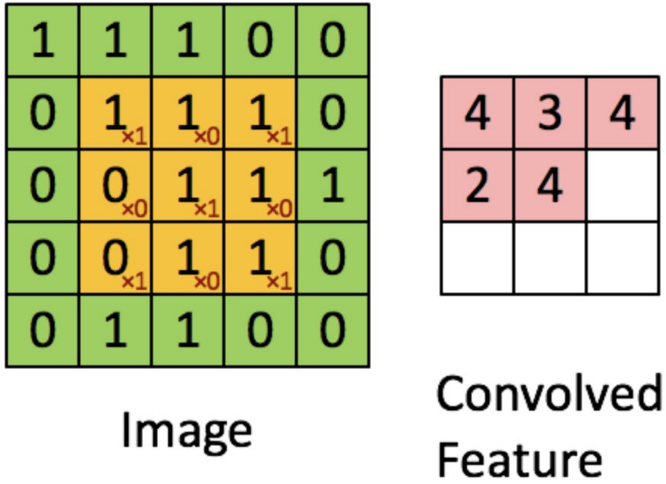
Answer: This is an existence theorem, i.e. it says that there is a network with 1 hidden layer that can do this (and doesn’t tell us how to find this network). However, the theorem doesn’t say that a network with 1 hidden layer is optimum in the sense of training time, ease of implementation, or (more importantly) generalization ability (i.e. ability to classify correctly new examples).

Exercise 5



3D Error Space

Exercise



1*1 + 1*0 + 1*1 + 0*0 + 1*1 + 1*0 + 0*1 + 1*0 + 1*1 = 4