Speech synthesis using Neural Networks

- what is a Neural Network?
- doing Text-to-Speech with a Neural Network
- training a Neural Network

What you should already know

- Text processing in the front end
  - what the available linguistic features are
  - how they can be **flattened** on to the phonetic sequence
  - how **categorical** linguistic features can be treated as **binary**
- HMM-based speech synthesis
  - how questions in a **regression tree** use those binary features
  - typical **speech parameters** used by vocoders

What you should already know

- Text processing in the front end
  - what the available linguistic features are
  - how they can be **flattened** on to the phonetic sequence
  - how **categorical** linguistic features can be treated as **binary**
- HMM-based speech synthesis
  - how questions in a **regression tree** use those binary features
  - typical **speech parameters** used by vocoders

What you should already know

- Text processing in the front end
  - what the available linguistic features are
  - how they can be **flattened** on to the phonetic sequence
  - how **categorical** linguistic features can be treated as **binary**
- HMM-based speech synthesis
  - how questions in a **regression tree** use those binary features
  - typical **speech parameters** used by vocoders

"**one-hot**" encoding

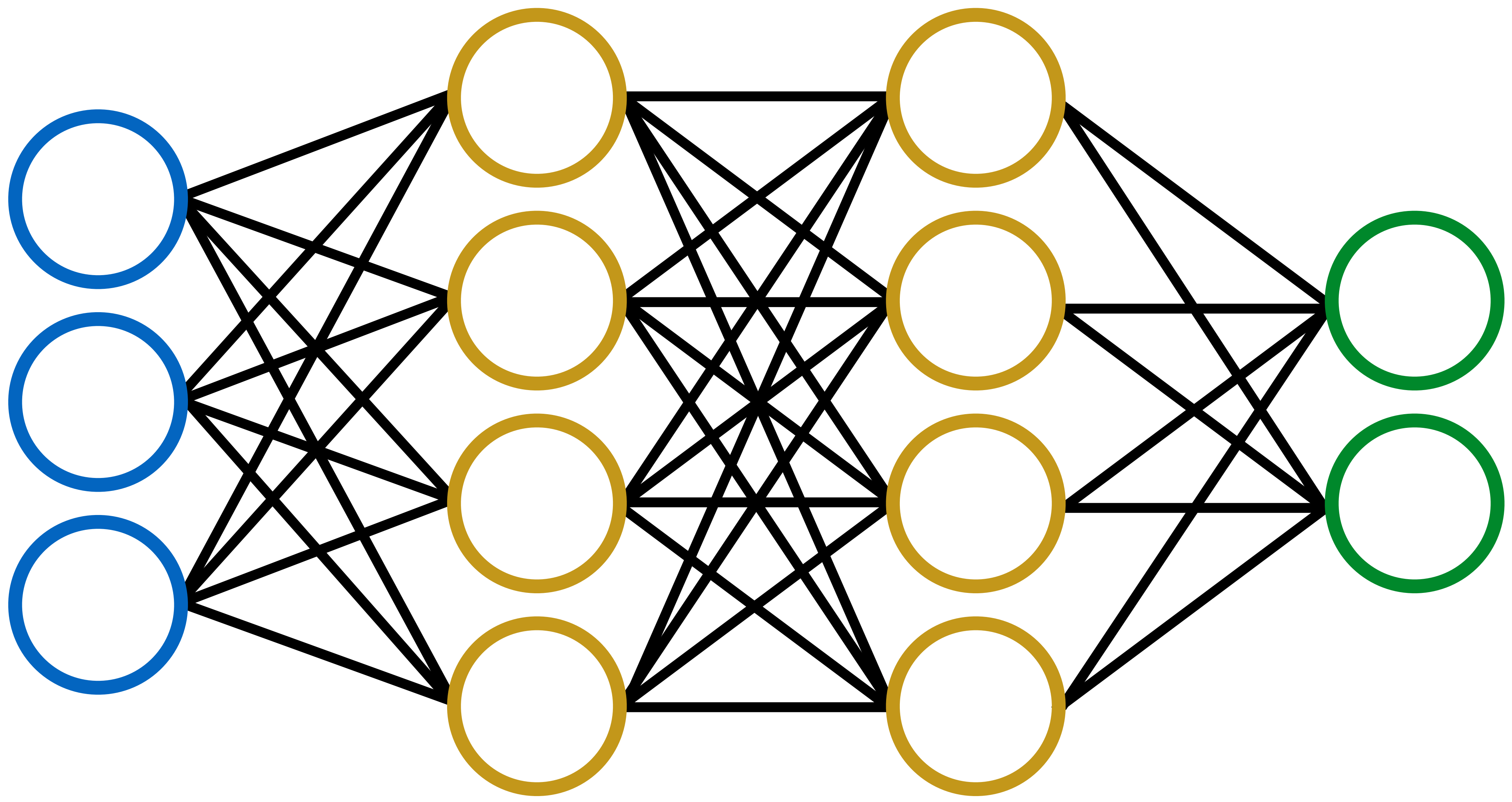
also known as

`1-of-K` or `1-of-N`

Speech synthesis using Neural Networks

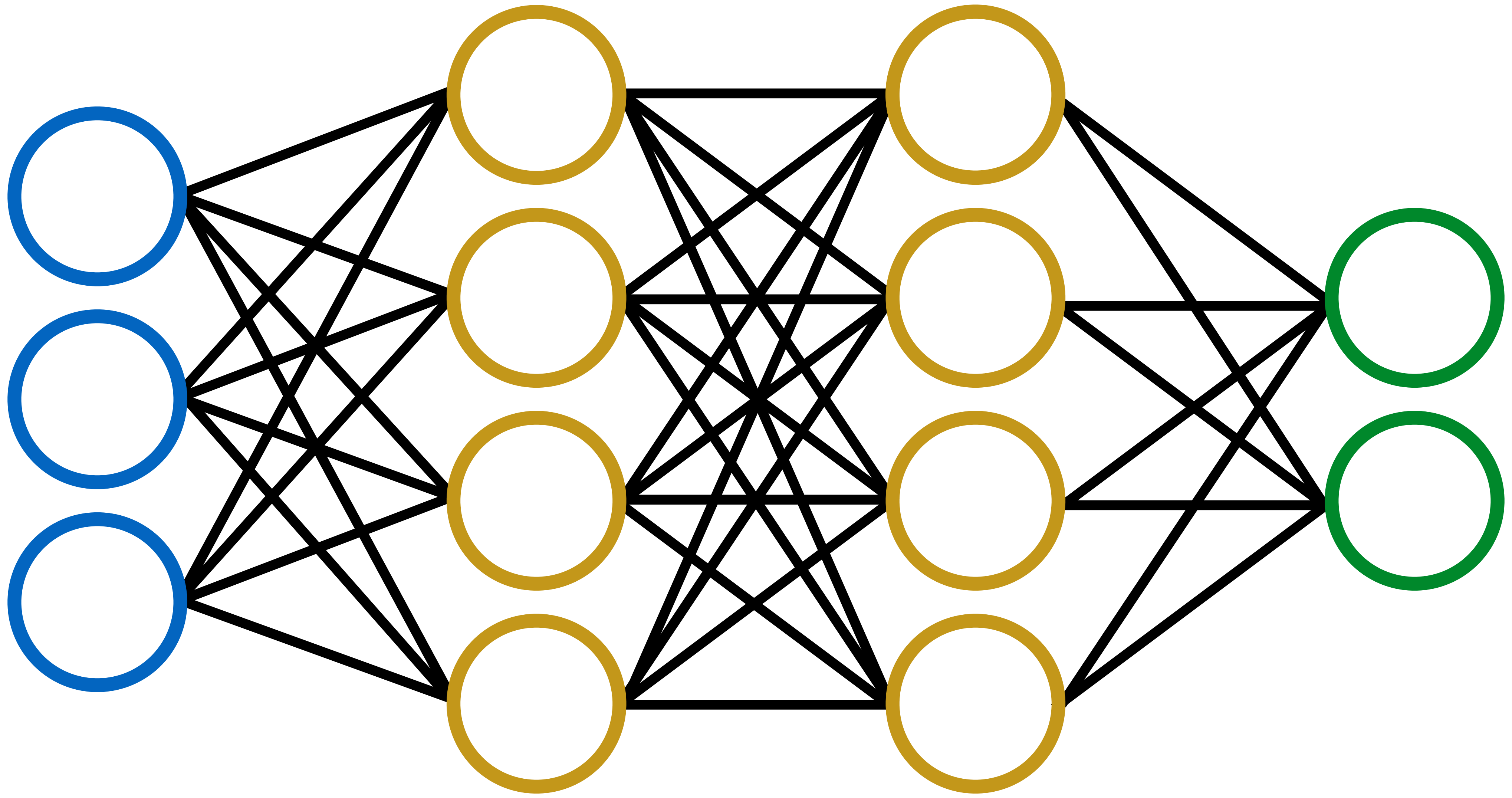
- <u>what is a Neural Network?</u>
- doing Text-to-Speech with a Neural Network
- training a Neural Network

A simple "feed forward" neural network

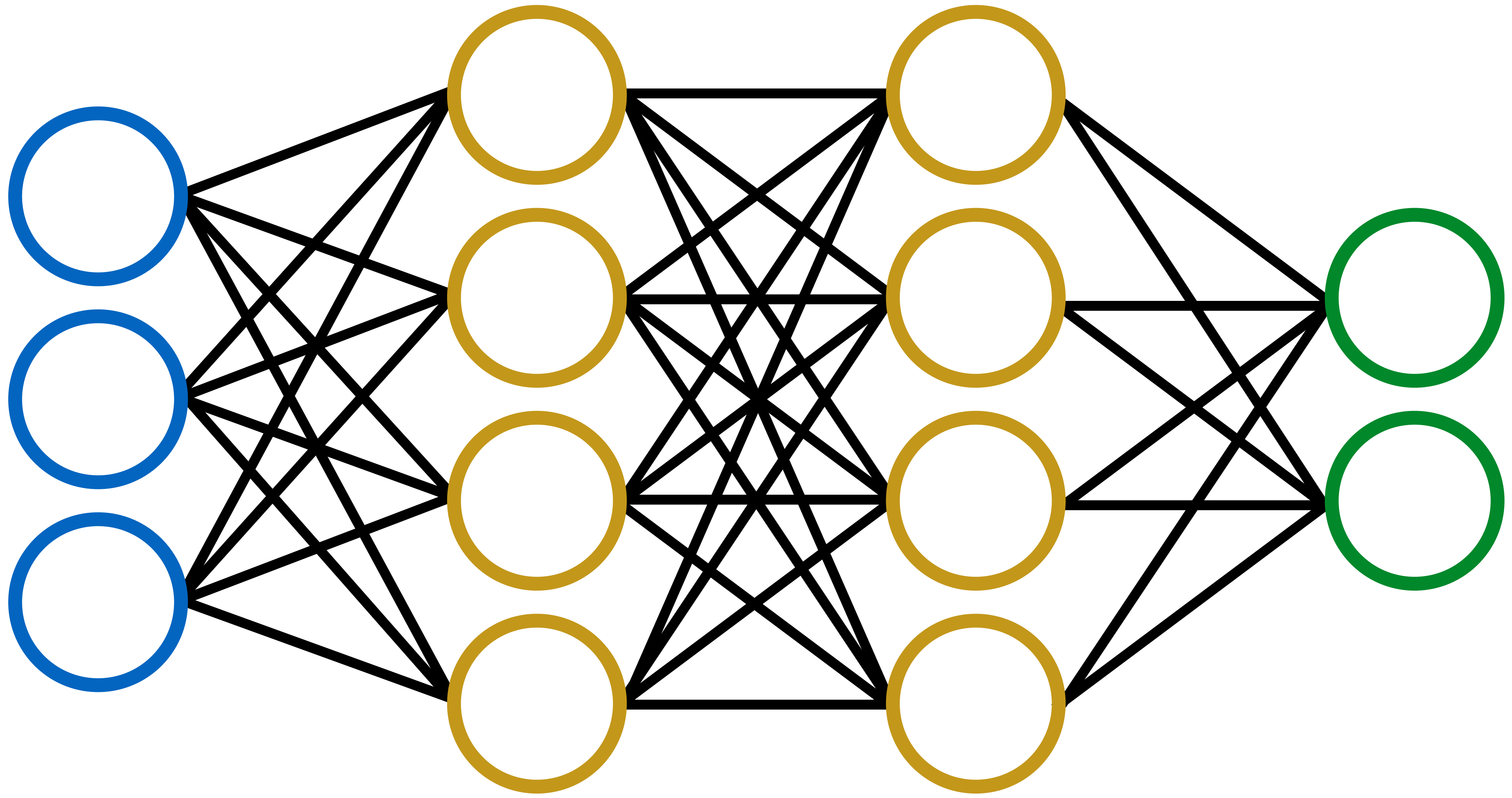


A simple "feed forward" neural network

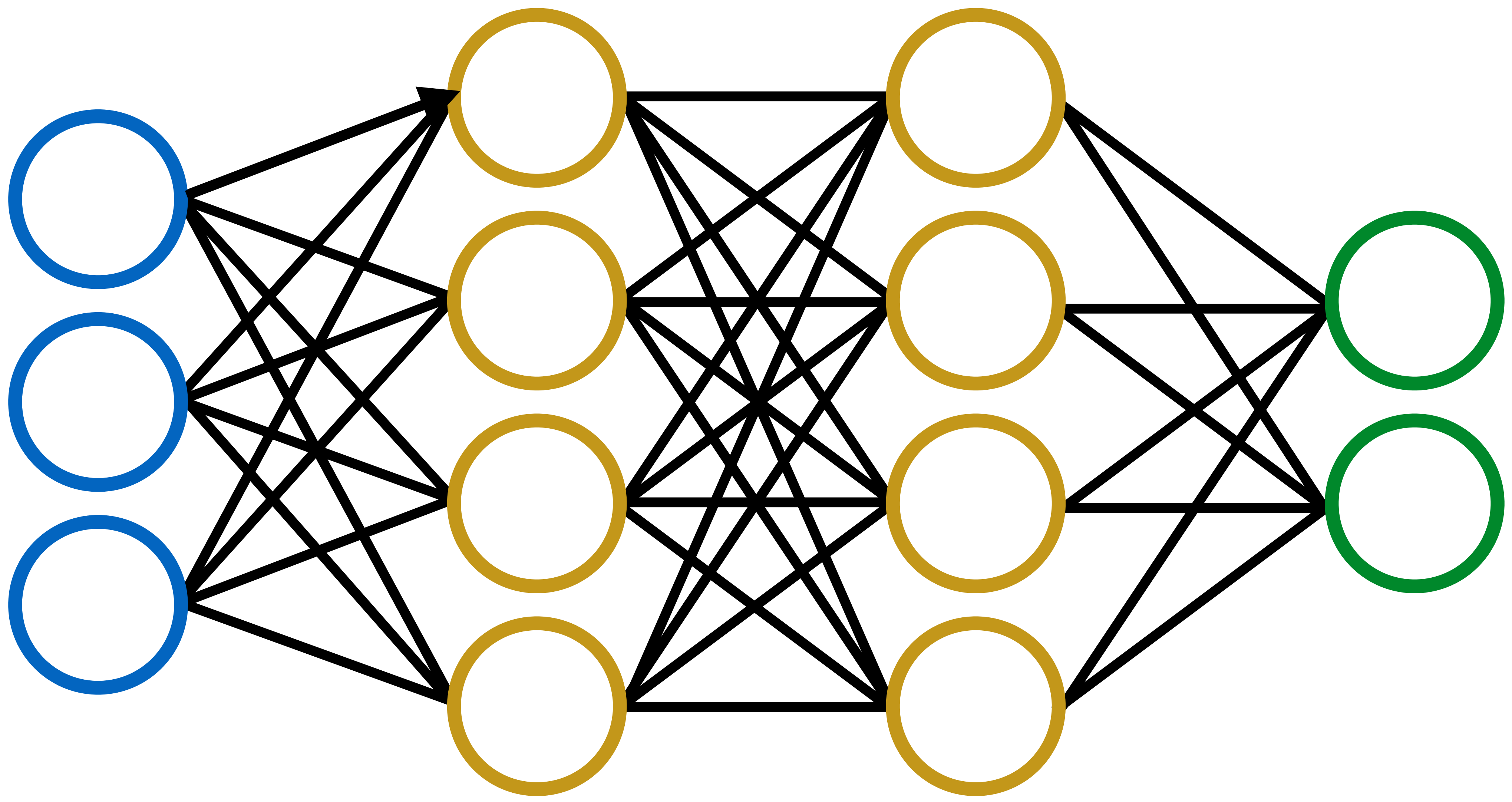
units (or "neurons"), each with an **activation function**



A simple "feed forward" neural network

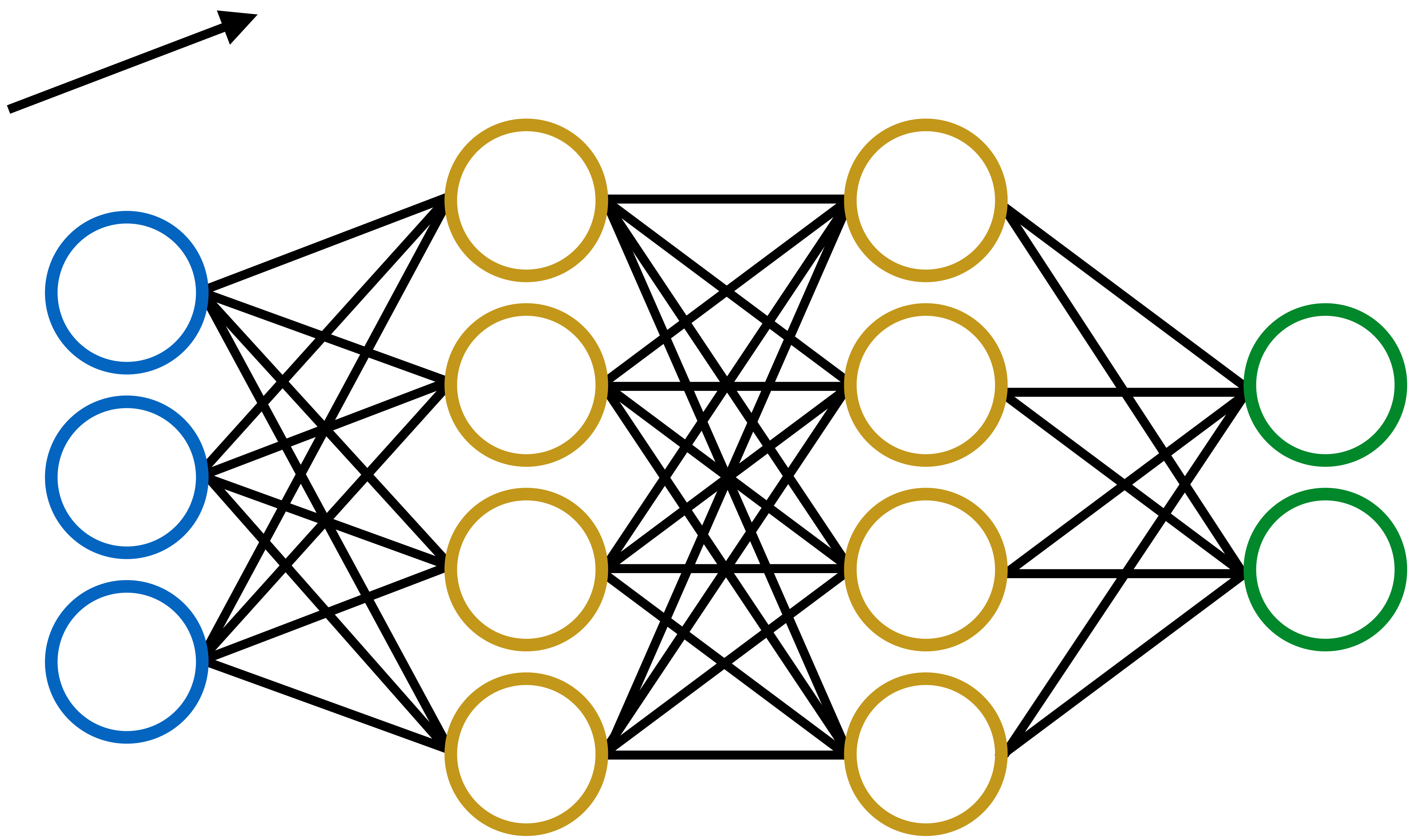


A simple "feed forward" neural network

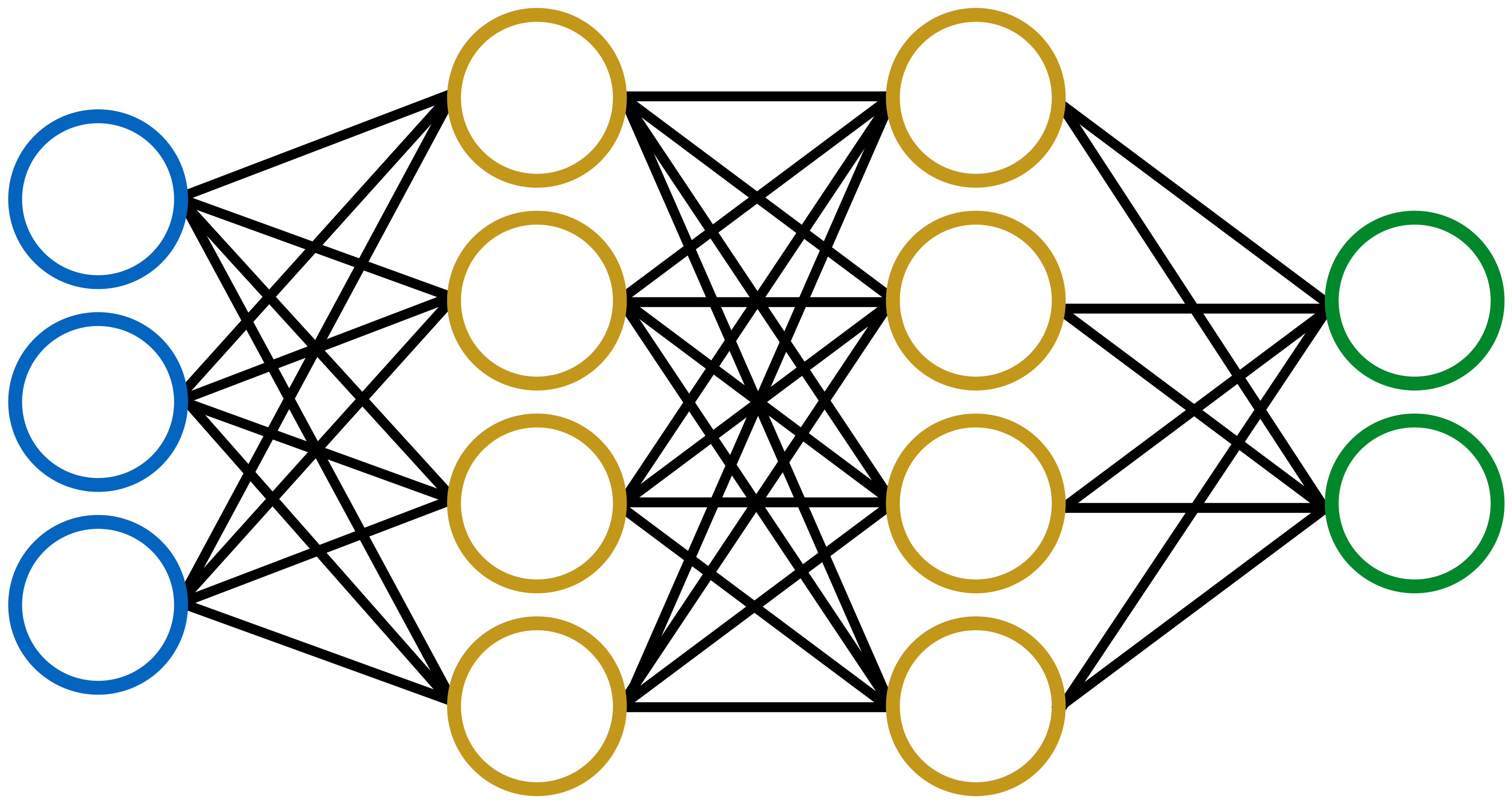


A simple "feed forward" neural network

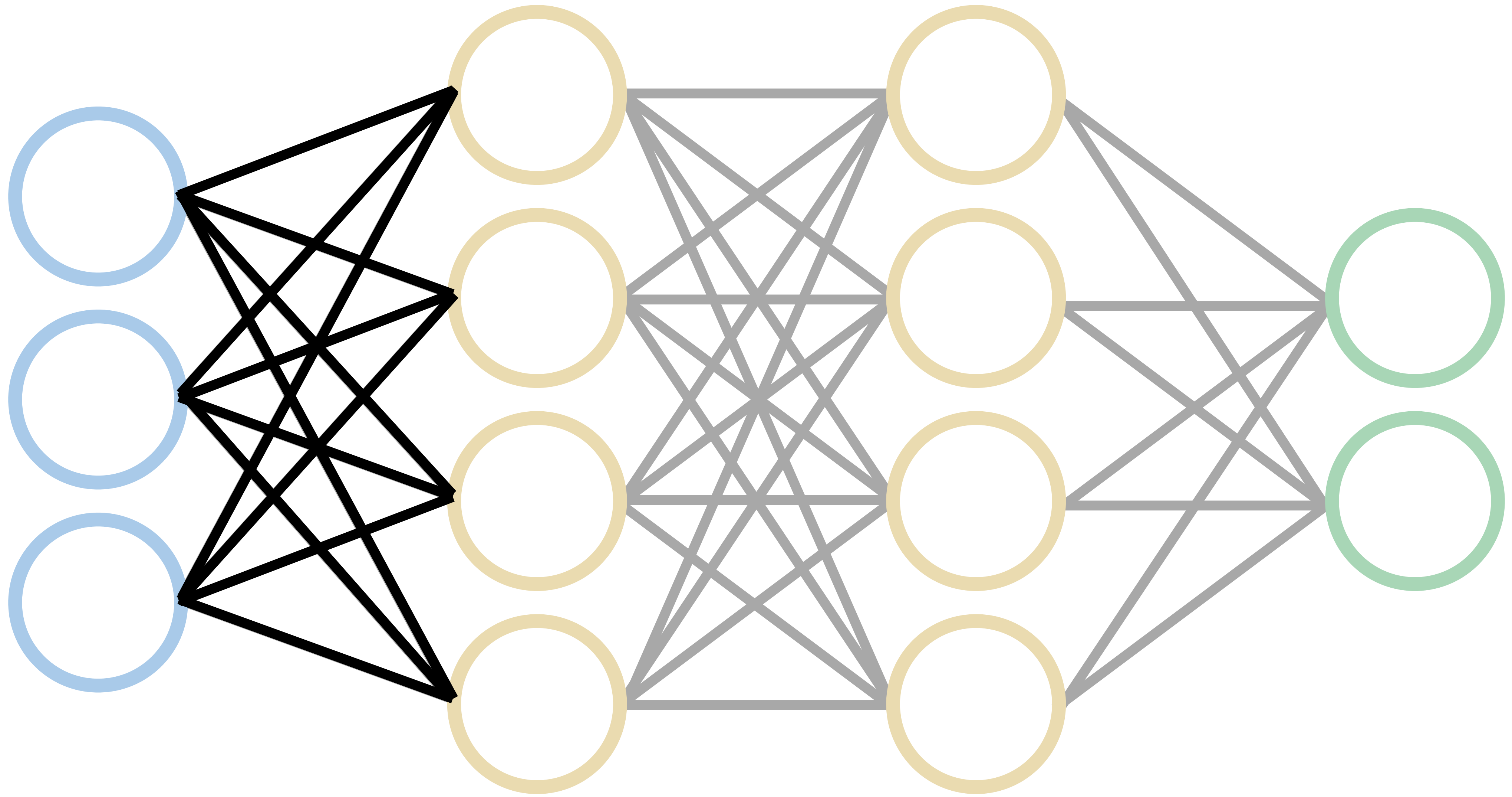
directed connections,
each with a **weight**



A simple "feed forward" neural network

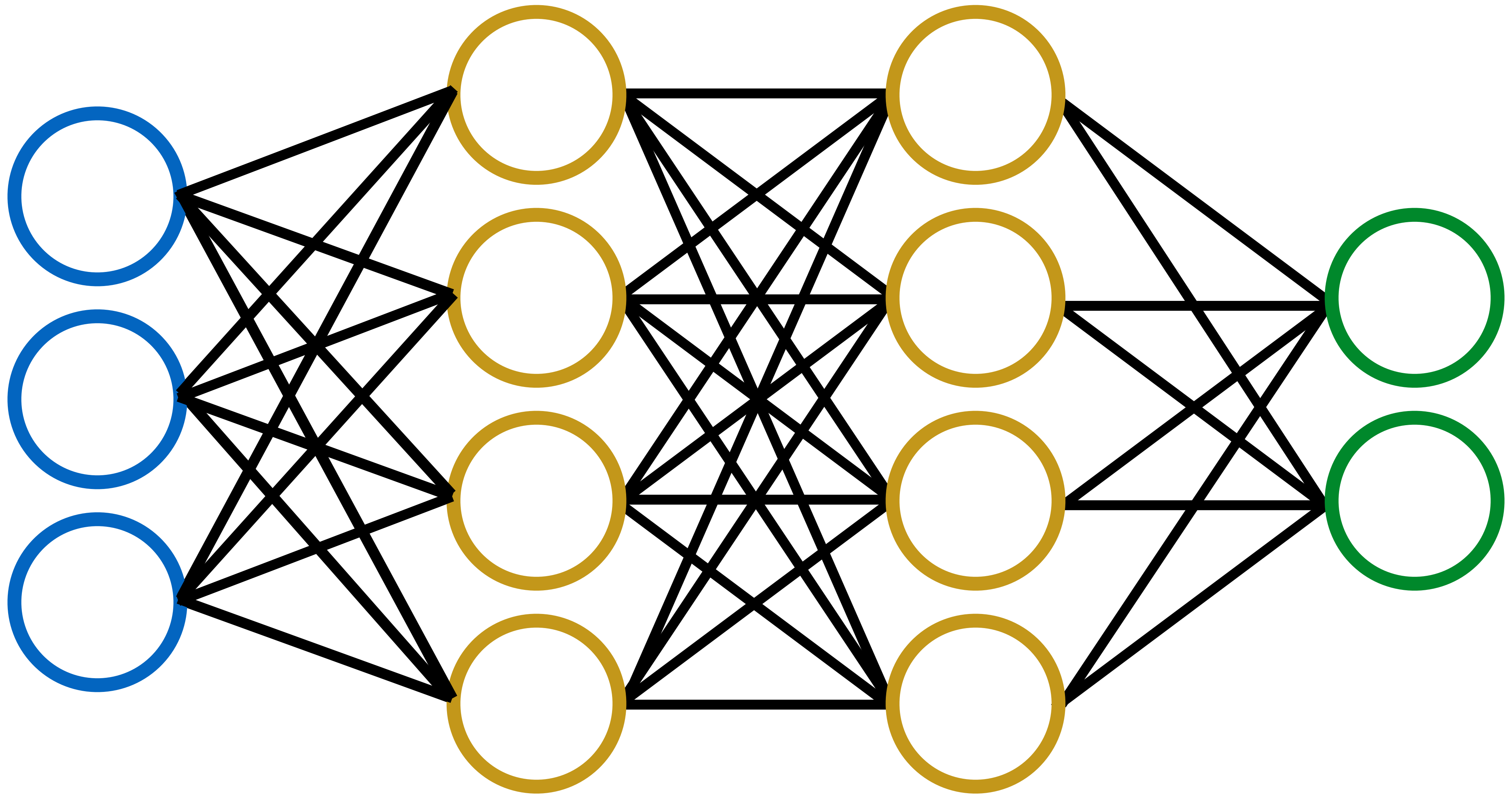


A simple "feed forward" neural network

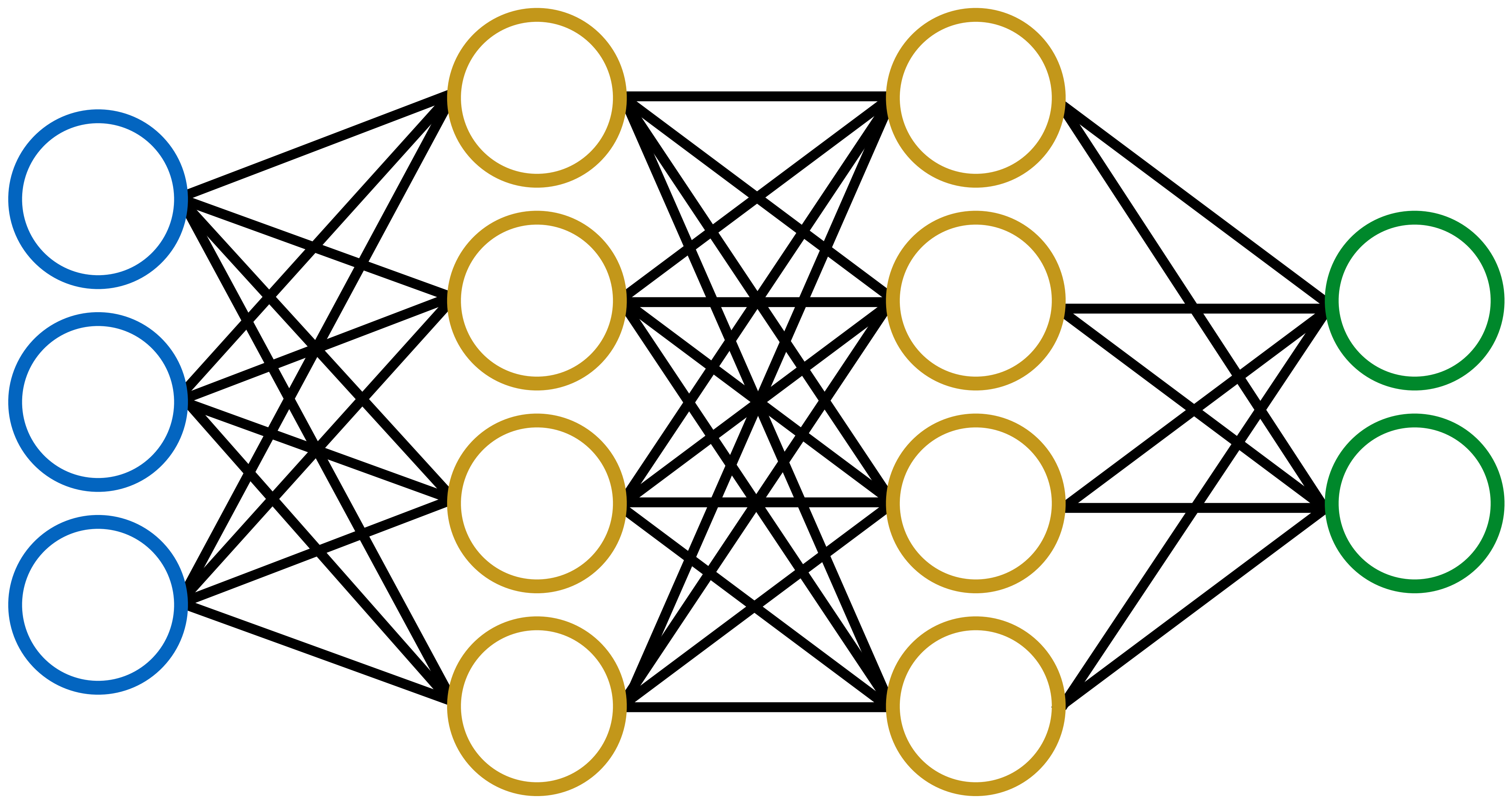


a weight **matrix**

A simple "feed forward" neural network

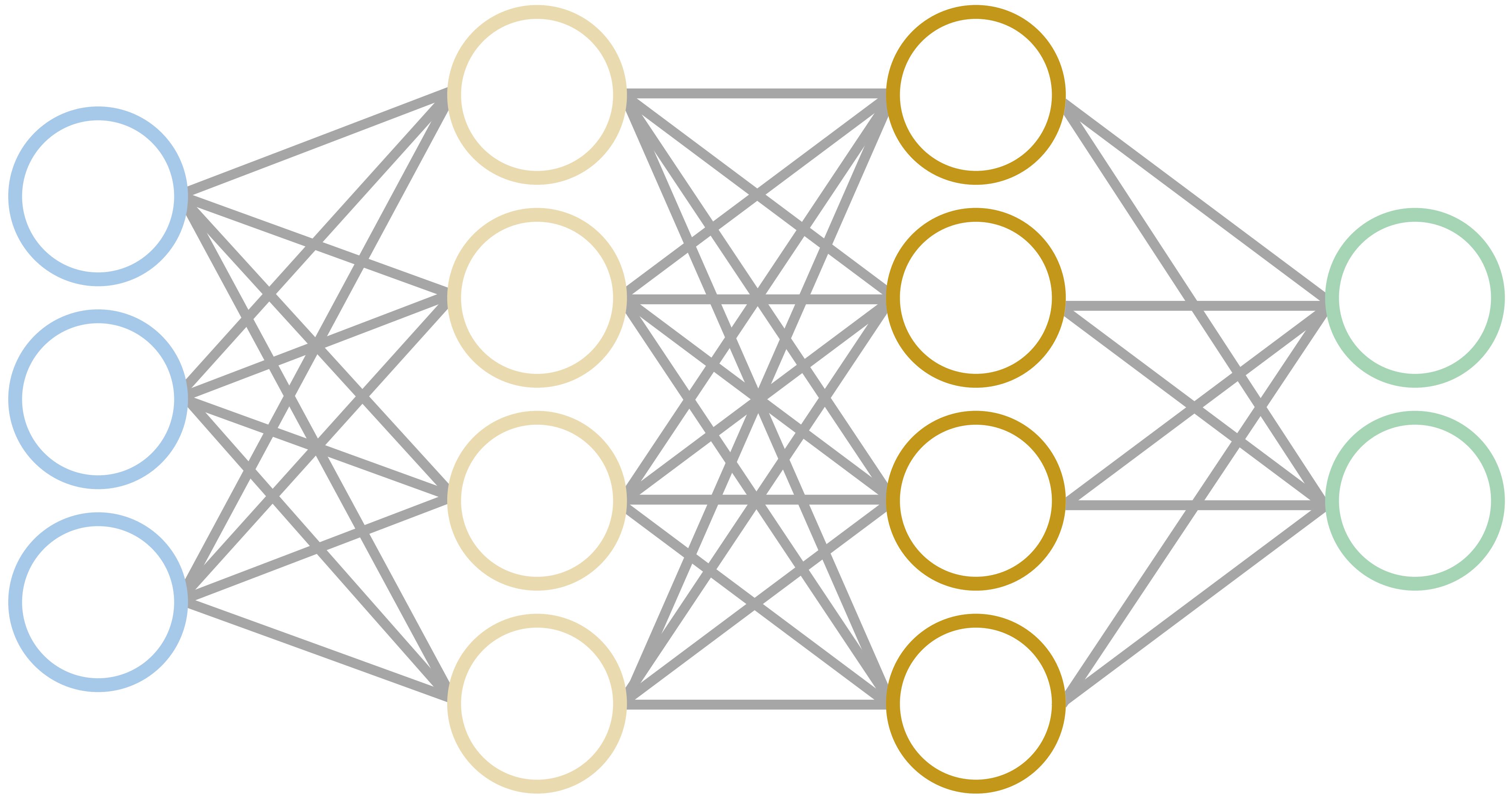


A simple "feed forward" neural network

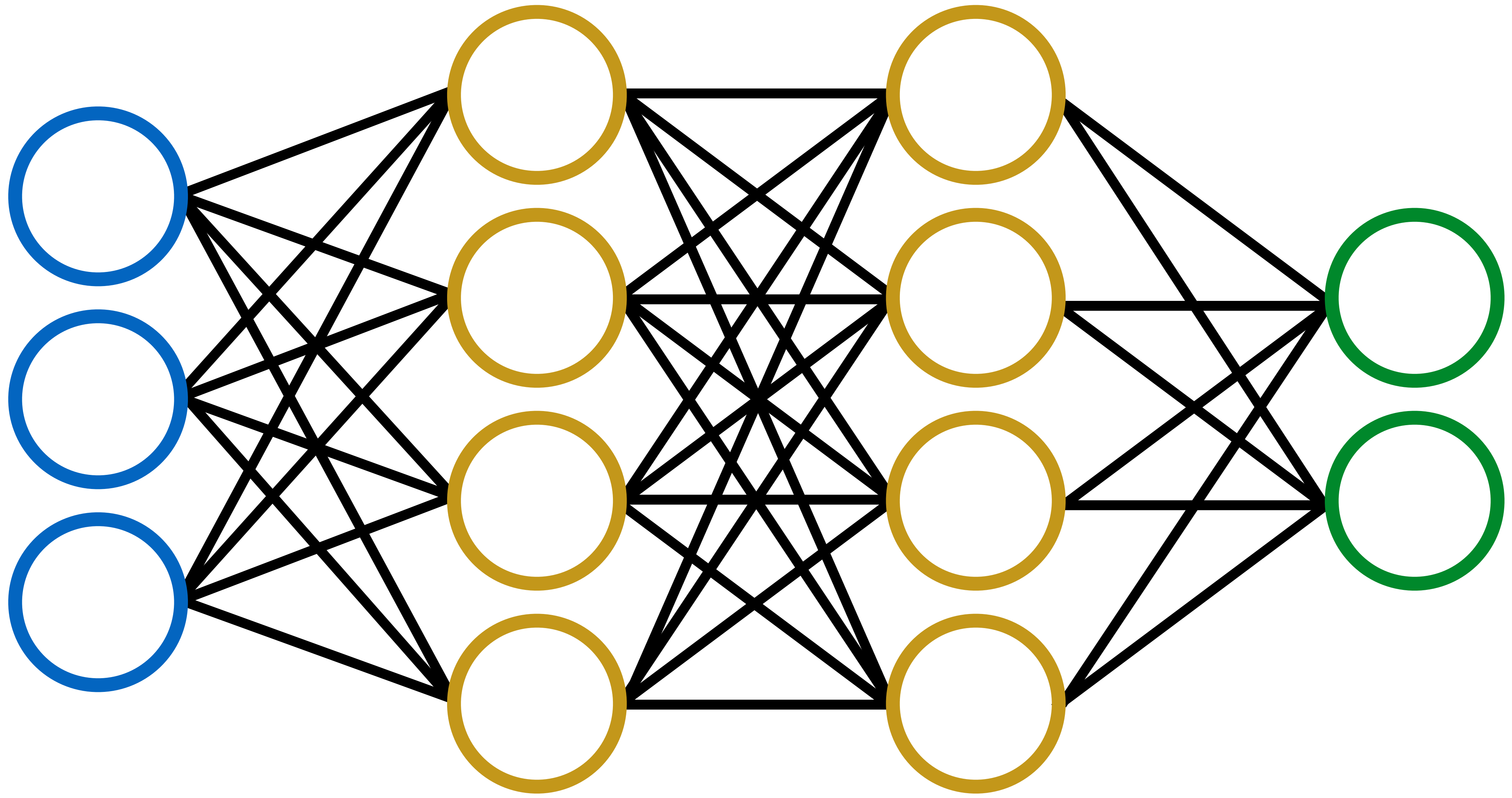


A simple "feed forward" neural network

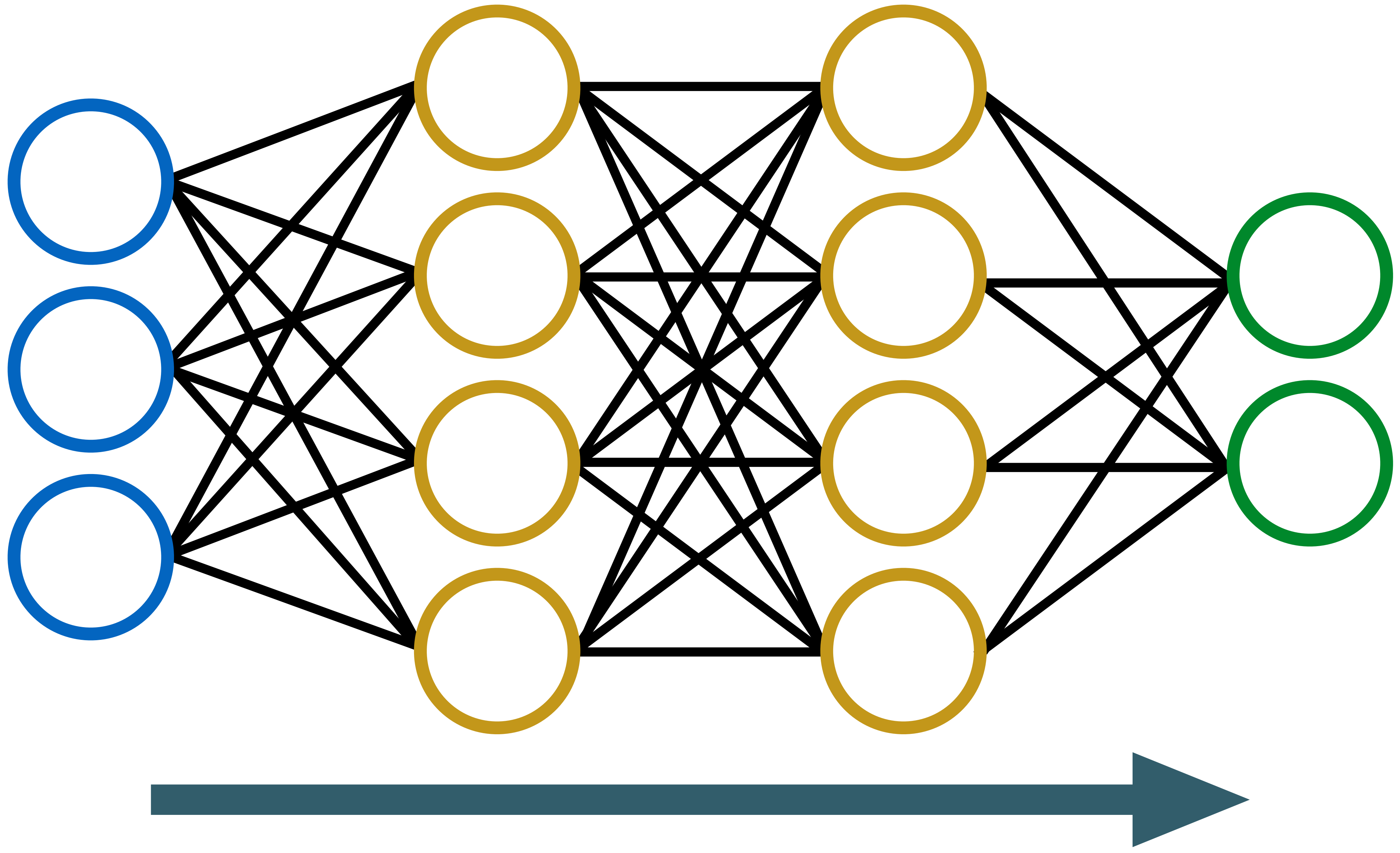
a hidden **layer**



A simple “feed forward” neural network

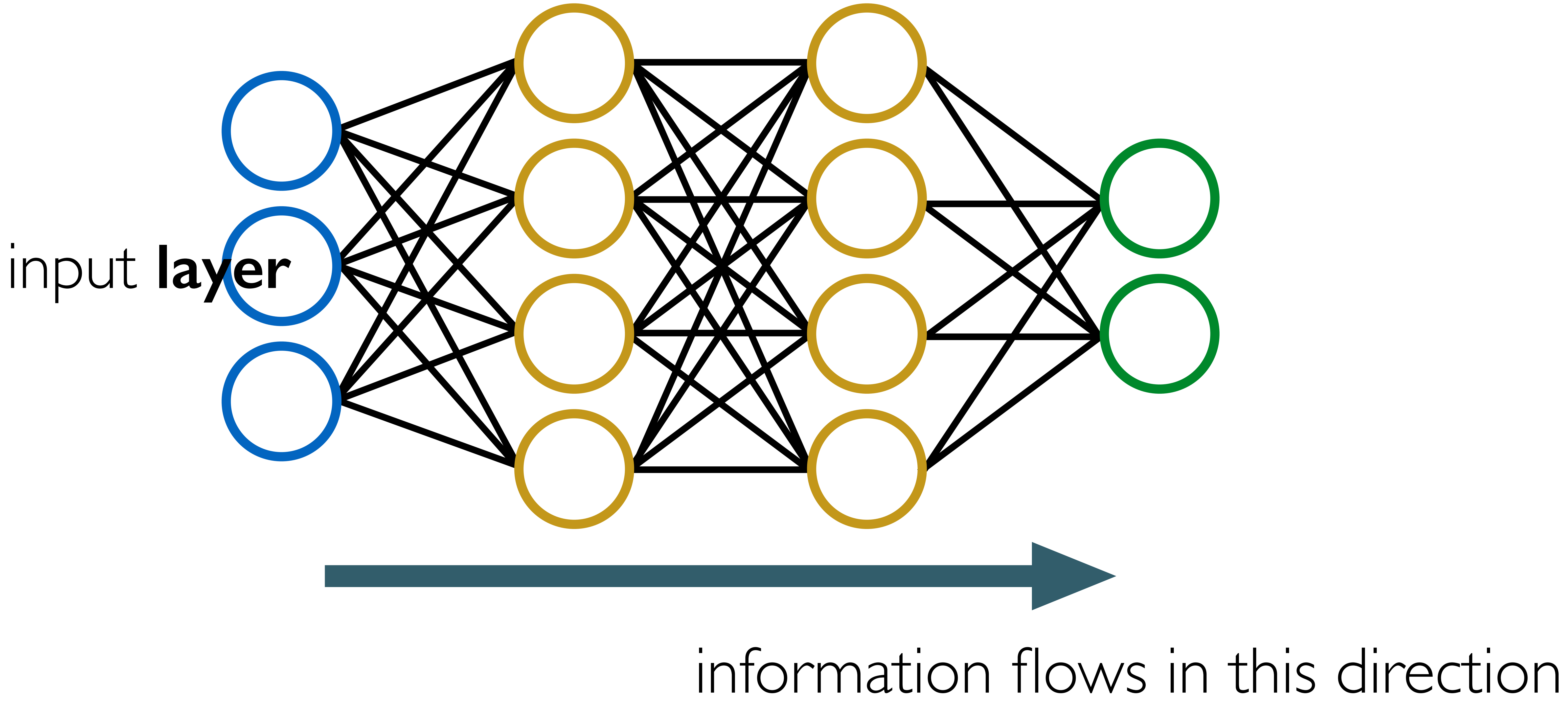


A simple "feed forward" neural network

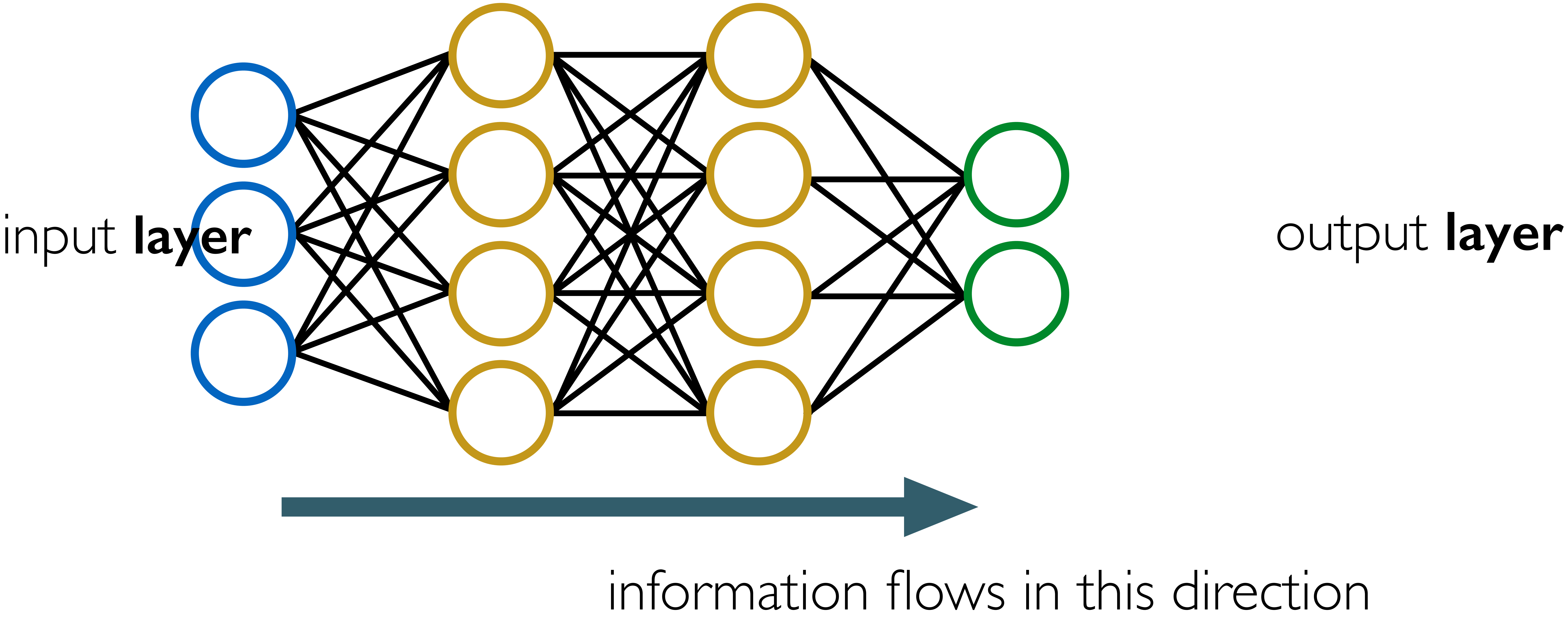


information flows in this direction

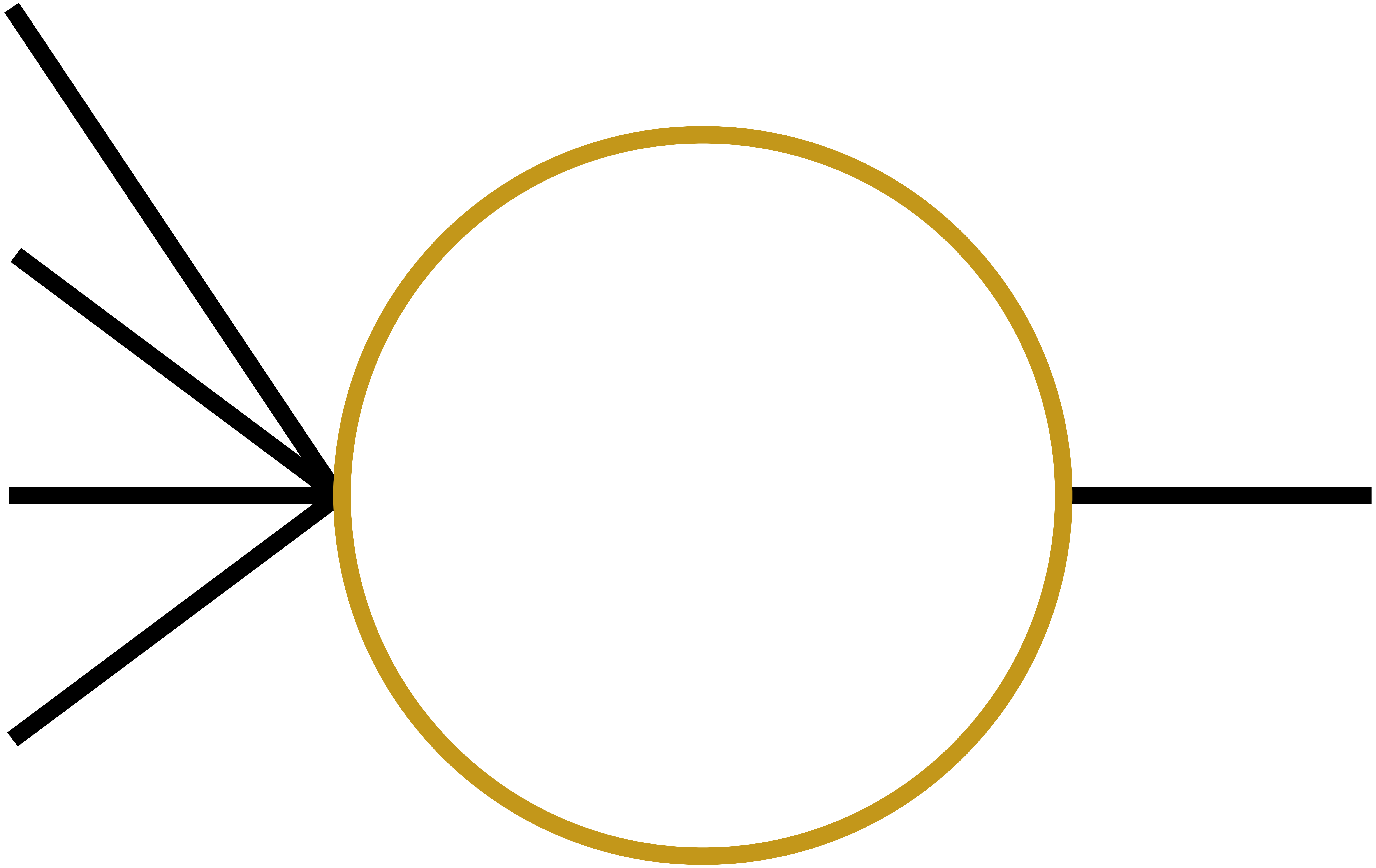
A simple "feed forward" neural network



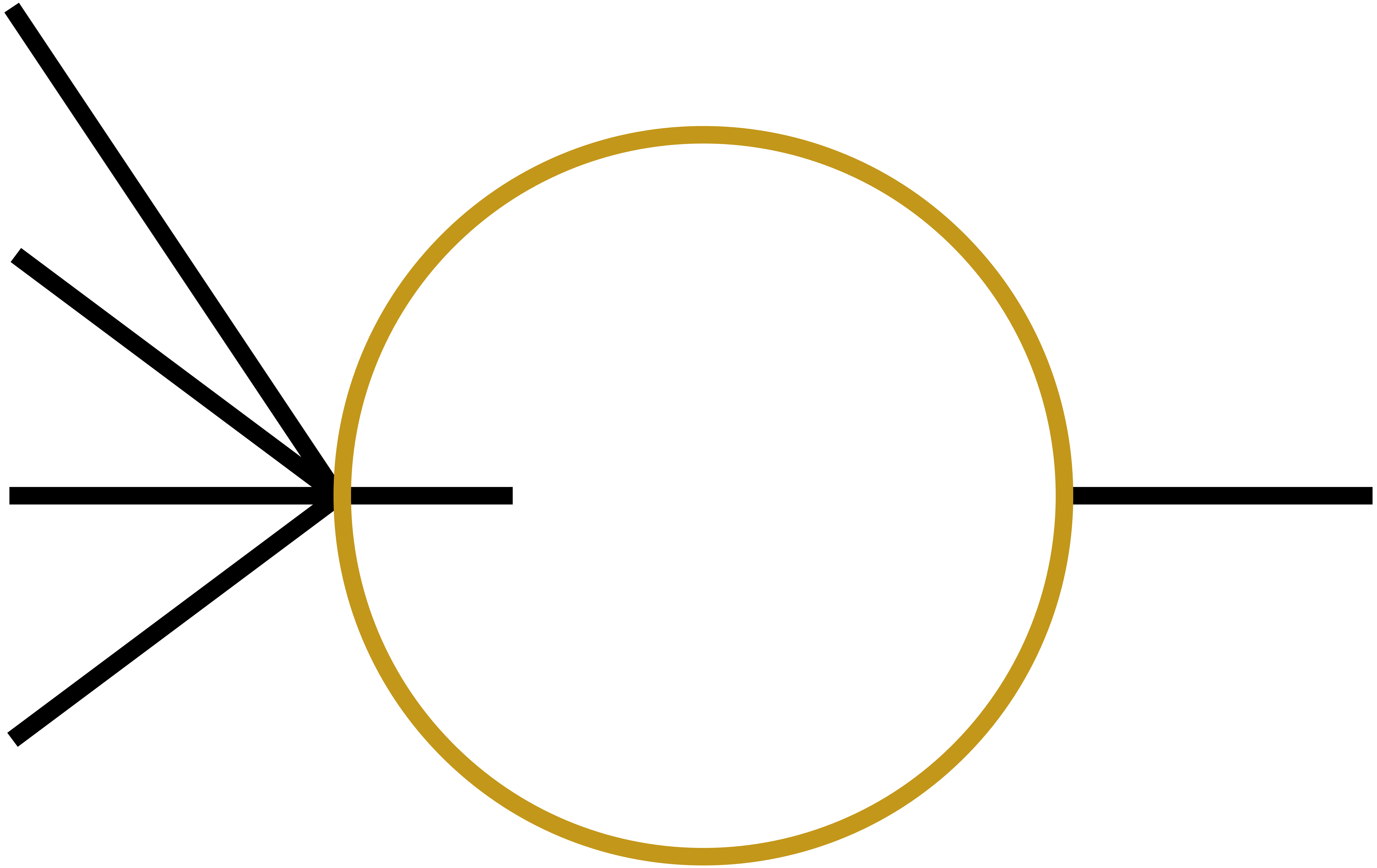
A simple "feed forward" neural network



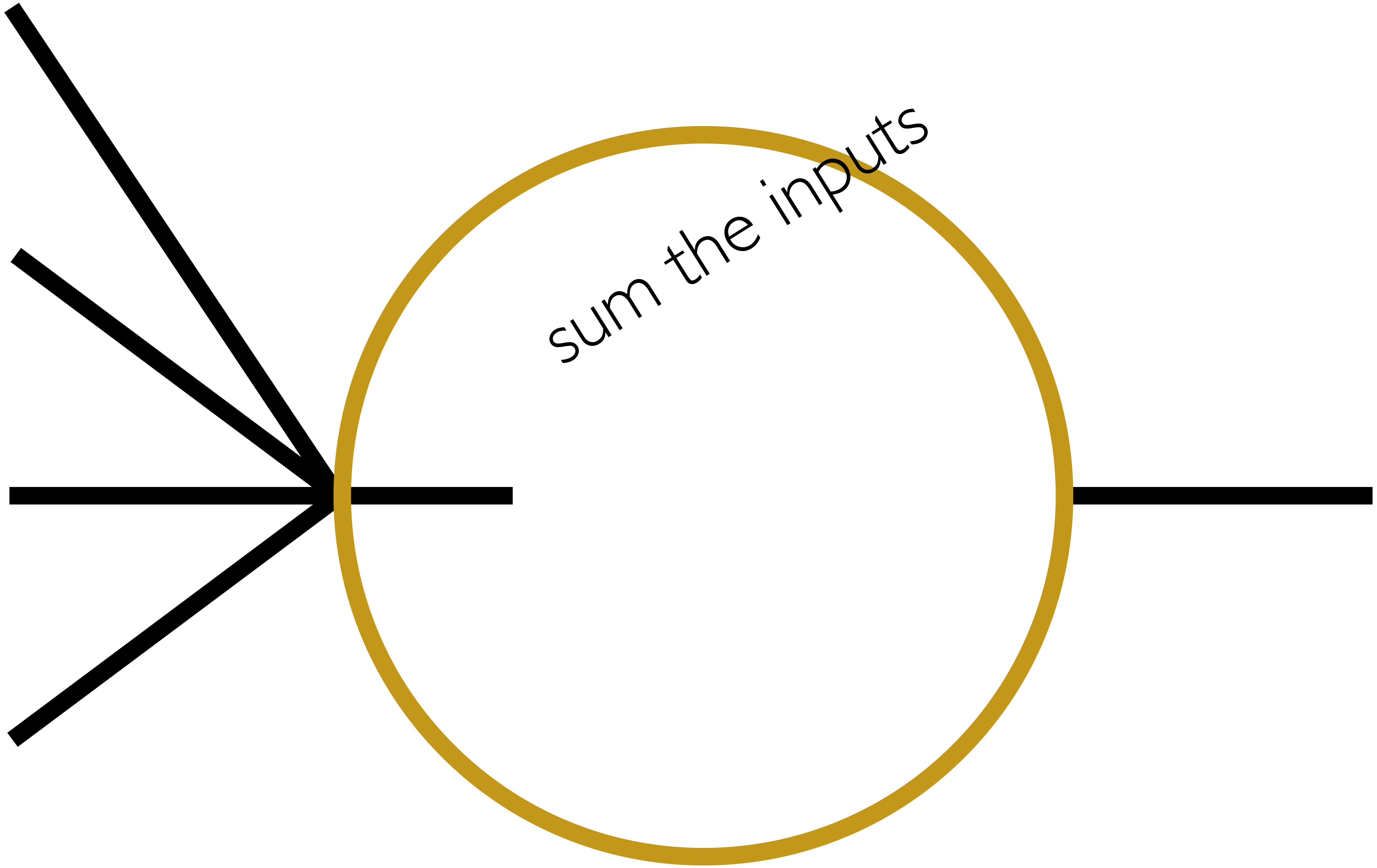
What is a unit, and what does it do?



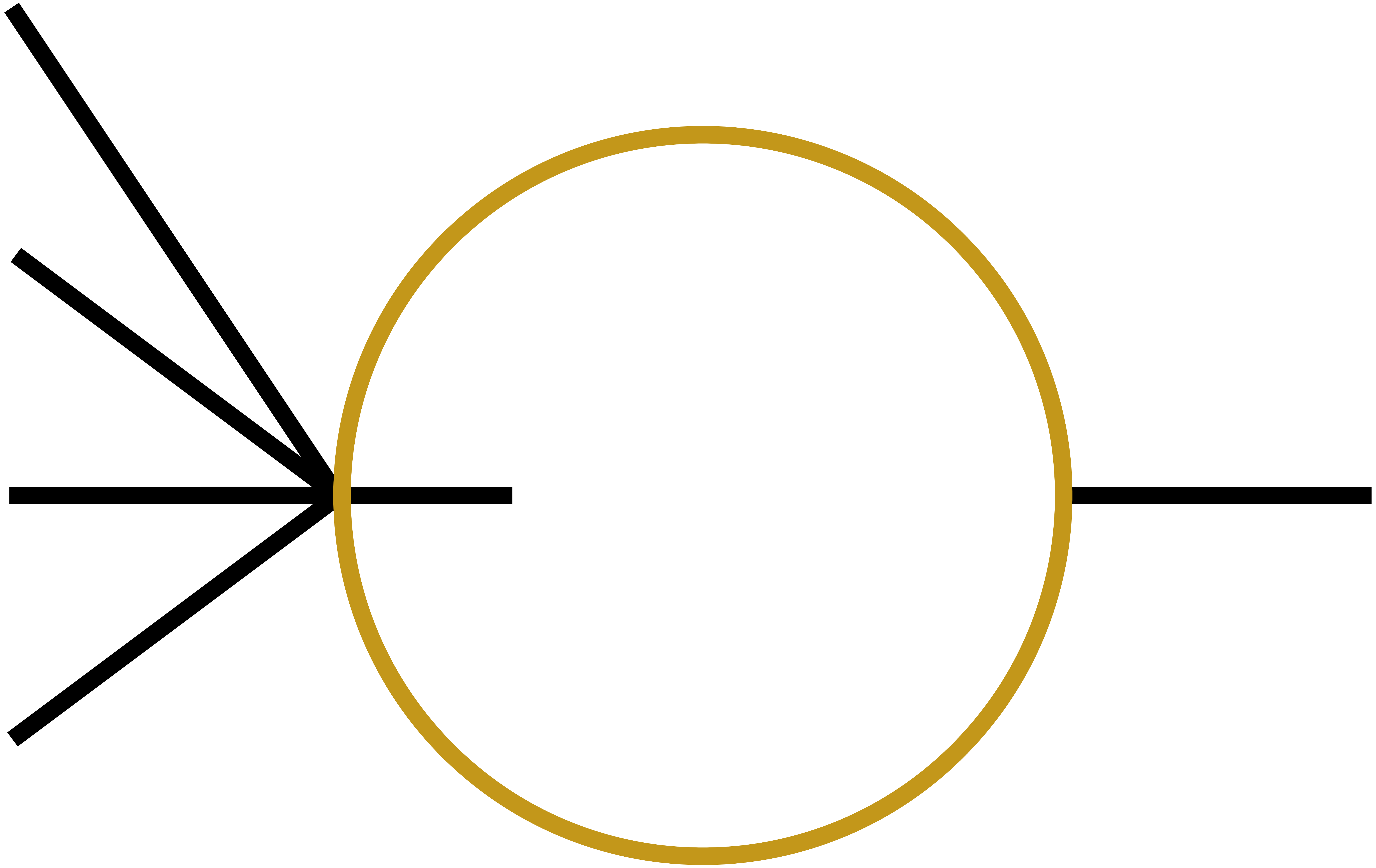
What is a unit, and what does it do?



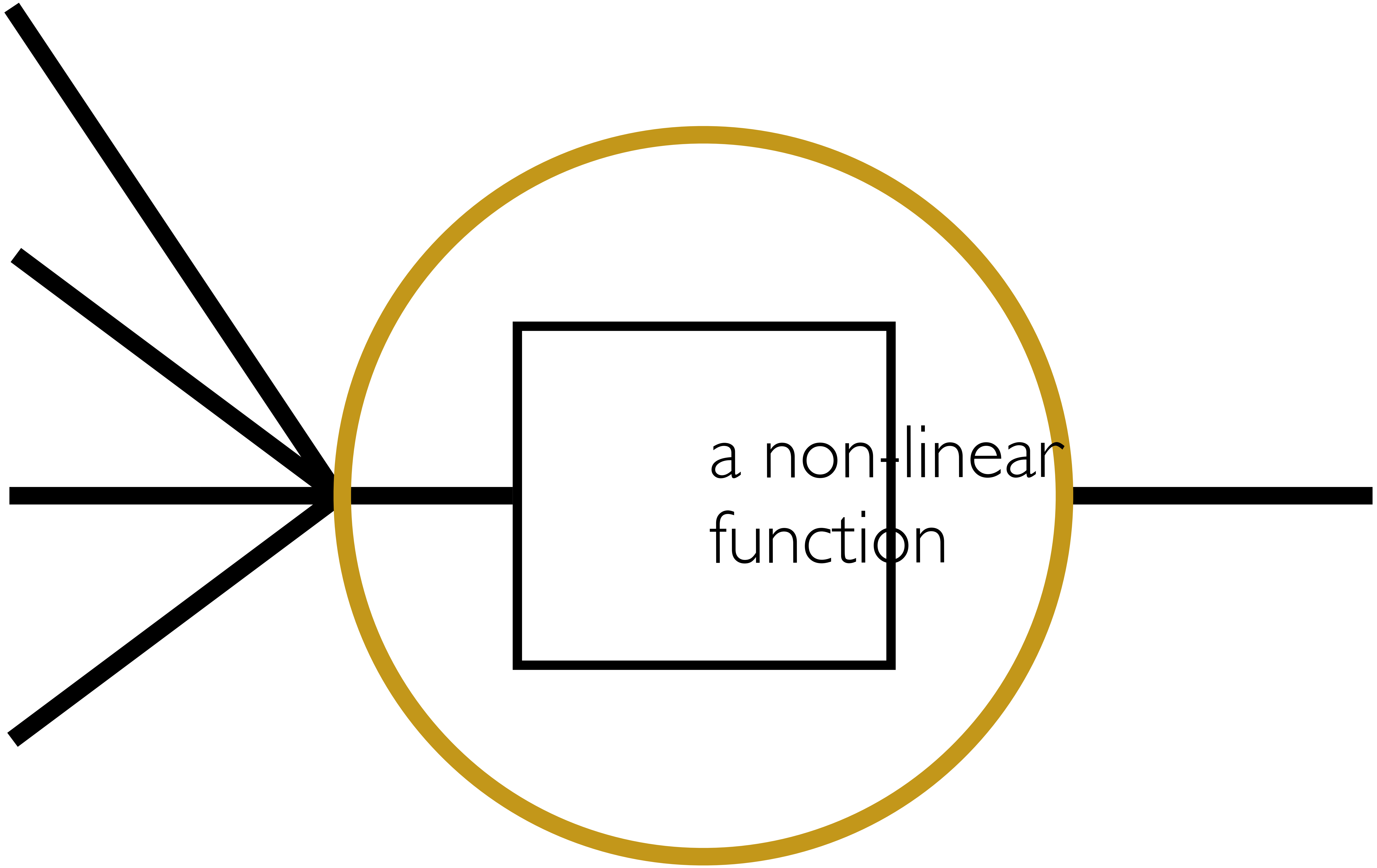
What is a unit, and what does it do?



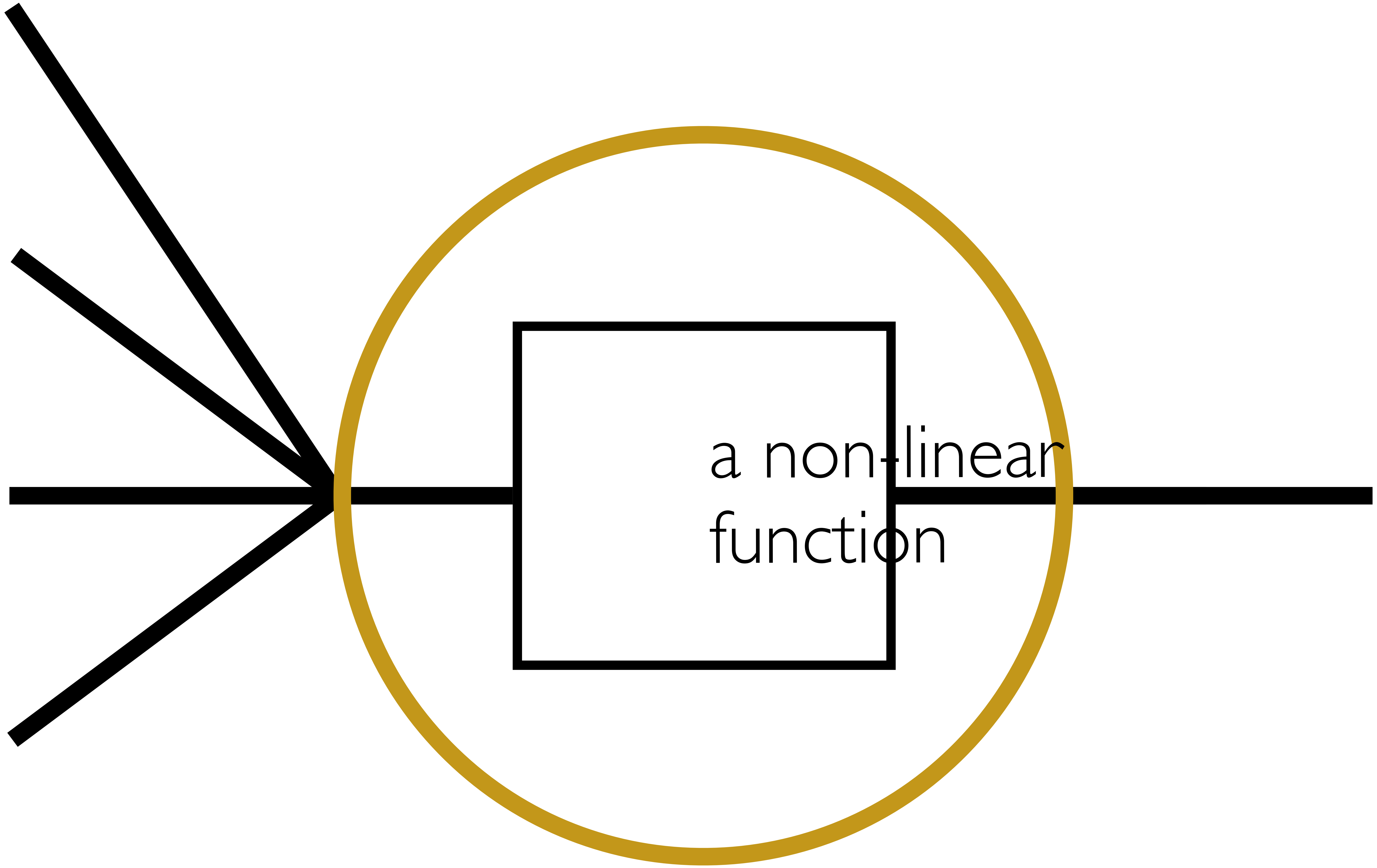
What is a unit, and what does it do?



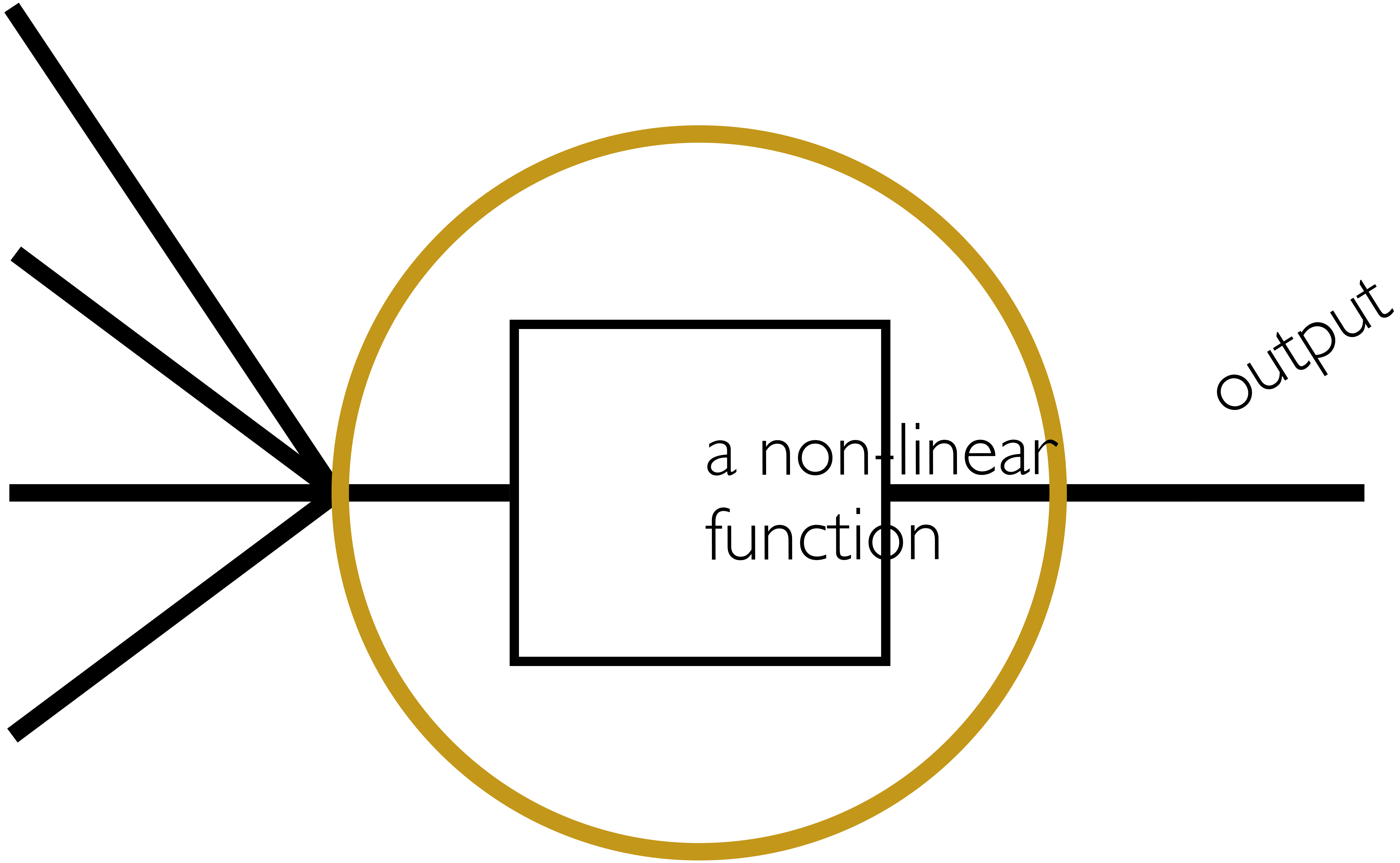
What is a unit, and what does it do?



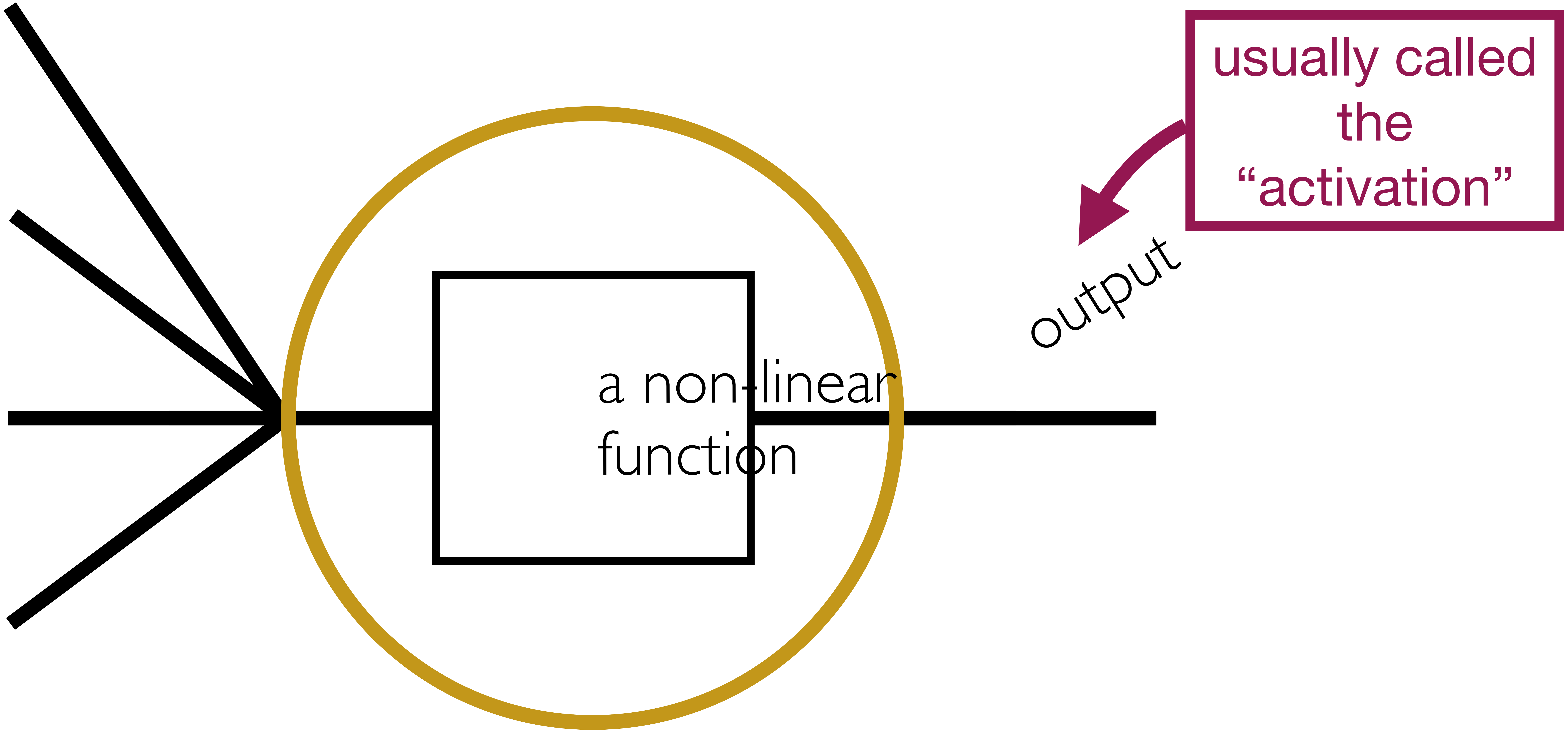
What is a unit, and what does it do?



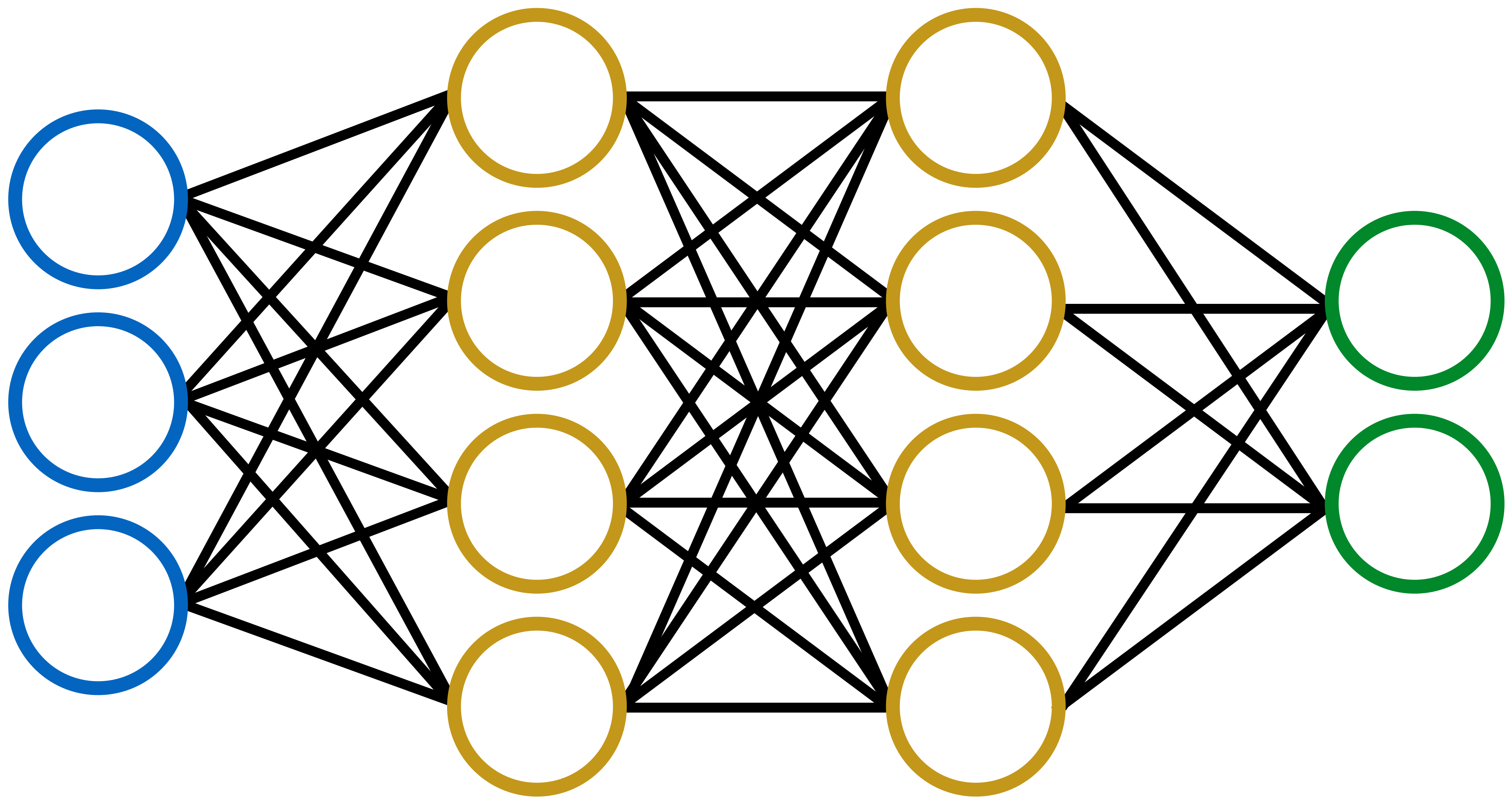
What is a unit, and what does it do?



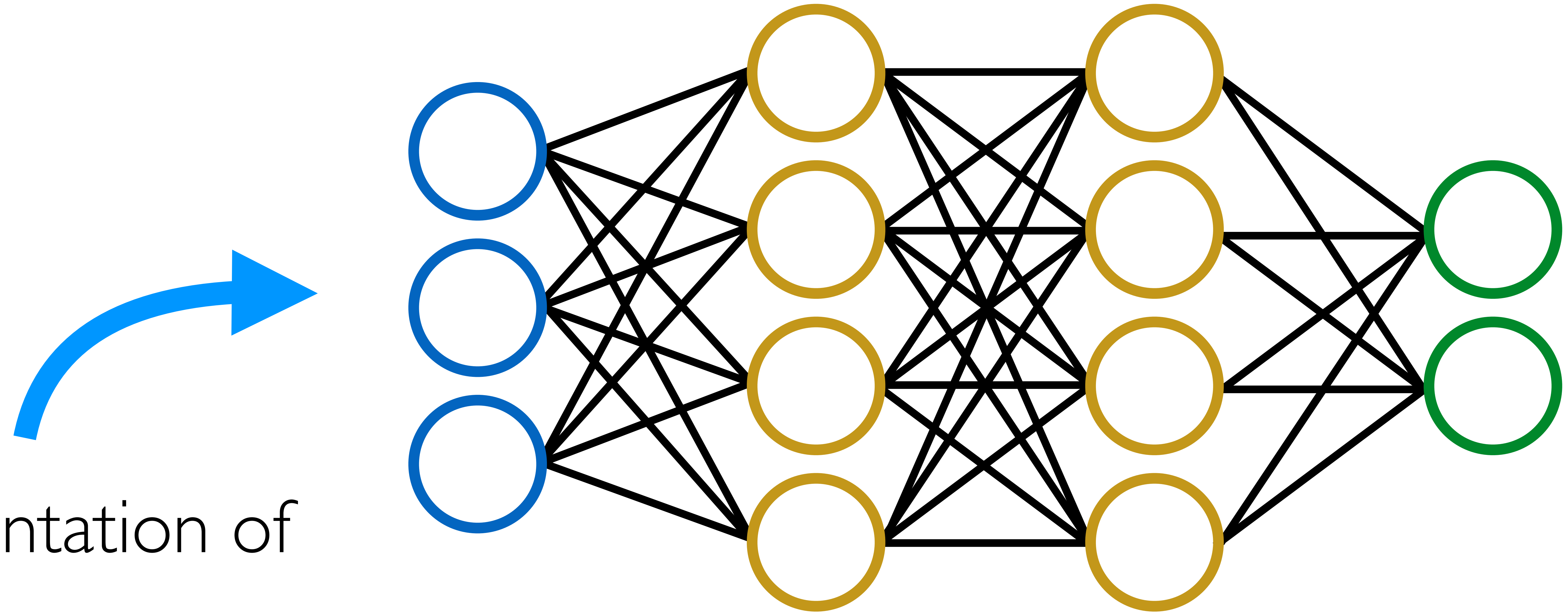
What is a unit, and what does it do?



What are all those layers for?

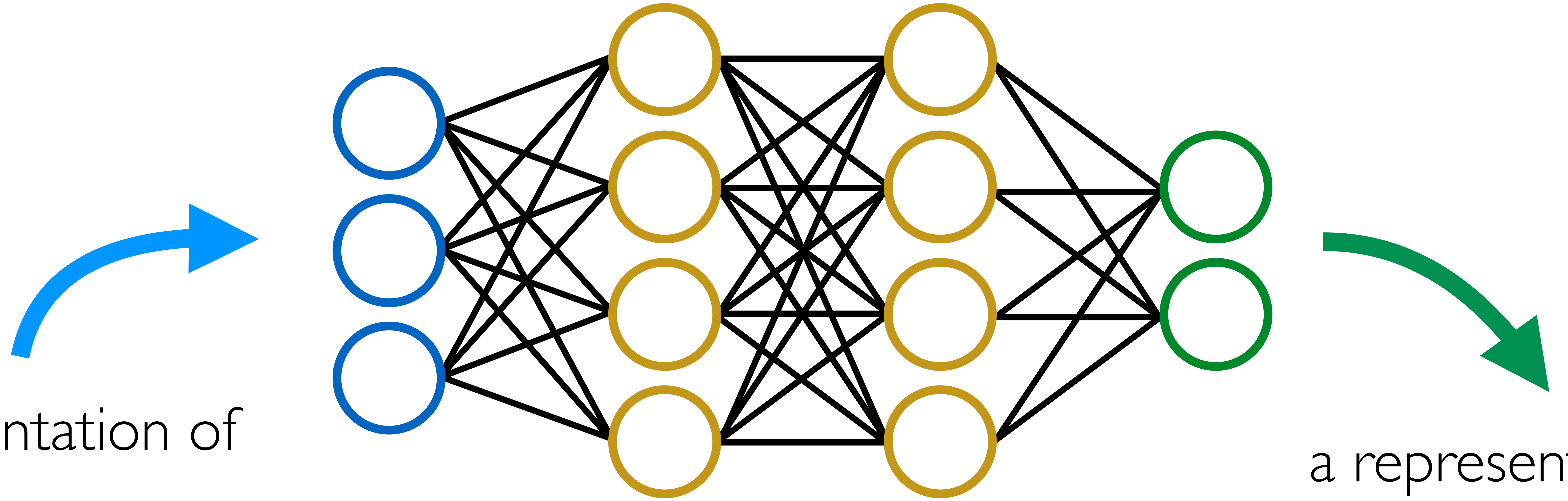


What are all those layers for?



a representation of
the input

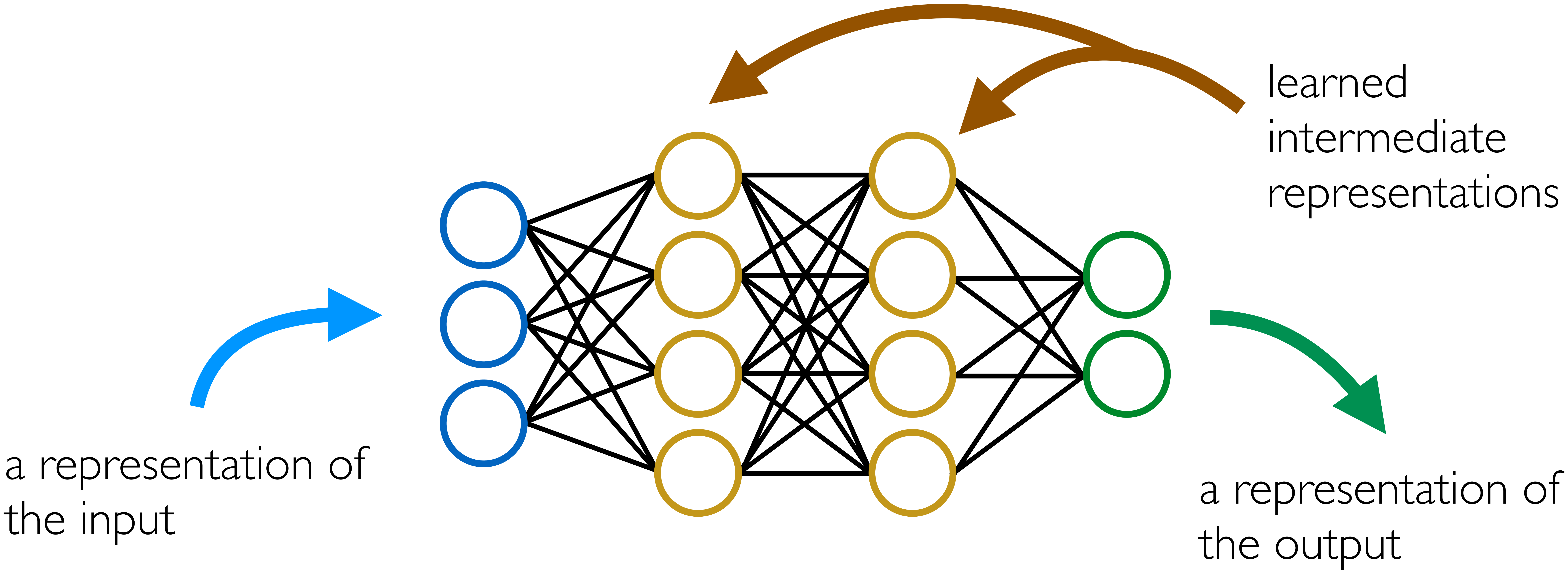
What are all those layers for?



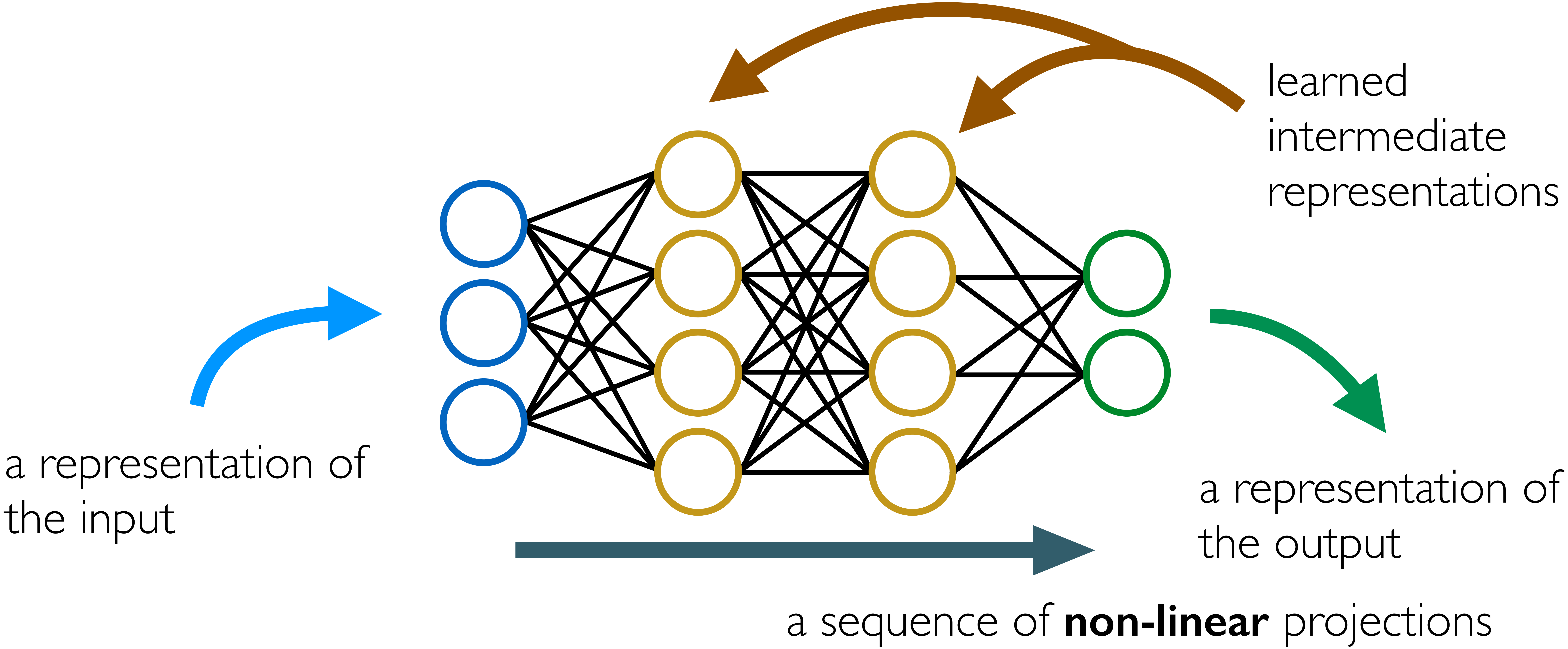
a representation of
the input

a representation of
the output

What are all those layers for?



What are all those layers for?



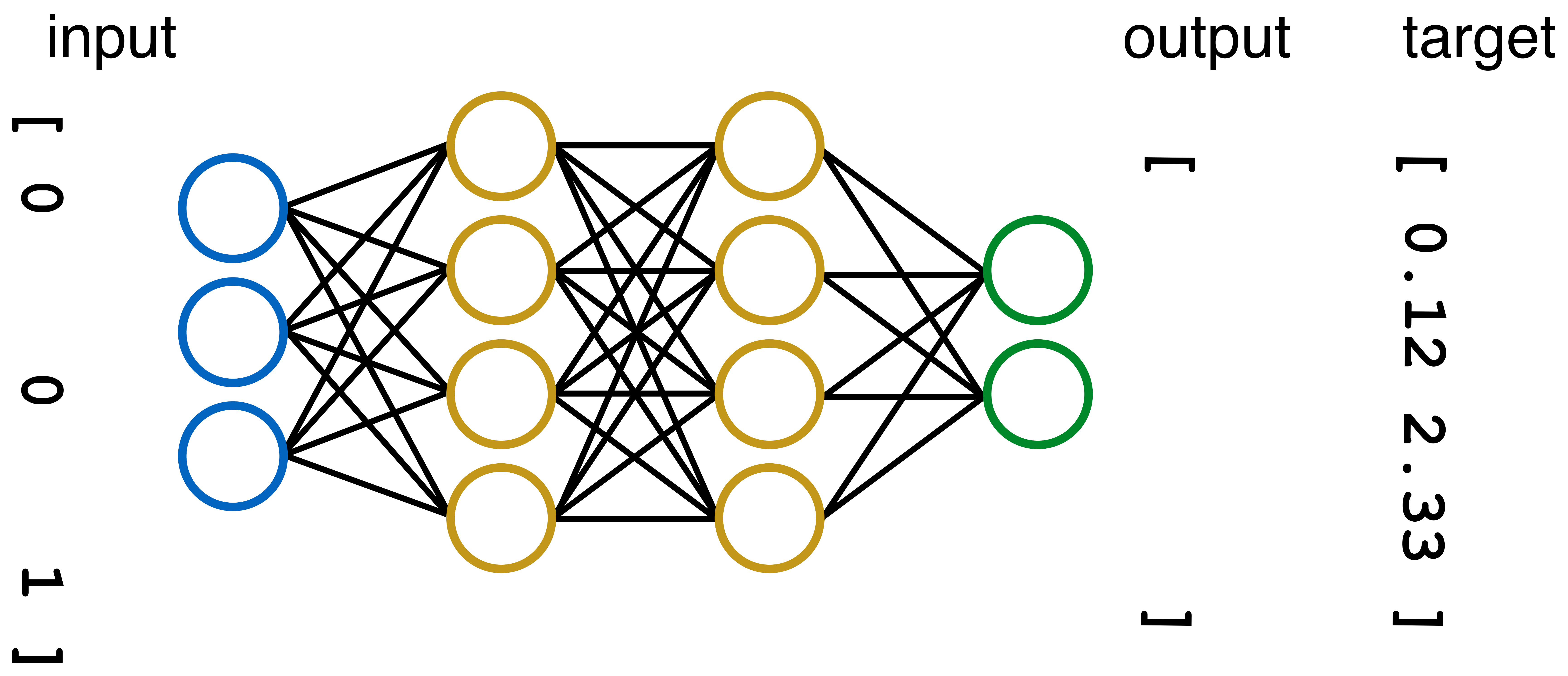
Training a neural network using back-propagation of the error

- what is the objective of training?
- notation
- taking the derivative
- deriving back-propagation

Supervised machine learning : input-output pairs (the output is the label for the input example)

```
[0 0 1 0 0 1 0 1 1 0 … 0.2  0.0]        [0.12 2.33 2.01 0.32 6.33 … ]
[0 0 1 0 0 1 0 1 1 0 … 0.2  0.1]        [0.43 2.11 1.99 0.39 4.83 … ]
…
[0 0 1 0 0 1 0 1 1 0 … 0.2  1.0]        [1.11 2.01 1.87 0.36 2.14 … ]
[0 0 1 0 0 1 0 1 1 0 … 0.4  0.0]        [1.52 1.82 1.89 0.34 1.04 … ]
[0 0 1 0 0 1 0 1 1 0 … 0.4  0.5]        [1.79 1.74 2.21 0.33 0.65 … ]
[0 0 1 0 0 1 0 1 1 0 … 0.4  1.0]        [1.65 1.58 2.68 0.31 0.73 … ]
…
[0 0 1 0 0 1 0 1 1 0 … 1.0  1.0]        [1.55 1.03 3.44 0.30 1.07 … ]
[0 0 0 1 1 1 0 1 0 0 … 0.2  0.0]        [1.92 0.99 3.89 0.29 1.45 … ]
[0 0 0 1 1 1 0 1 0 0 … 0.2  0.2]        [2.38 1.13 4.02 0.28 1.98 … ]
[0 0 0 1 1 1 0 1 0 0 … 0.2  0.4]        [2.65 1.98 3.94 0.29 2.16 … ]
…
```

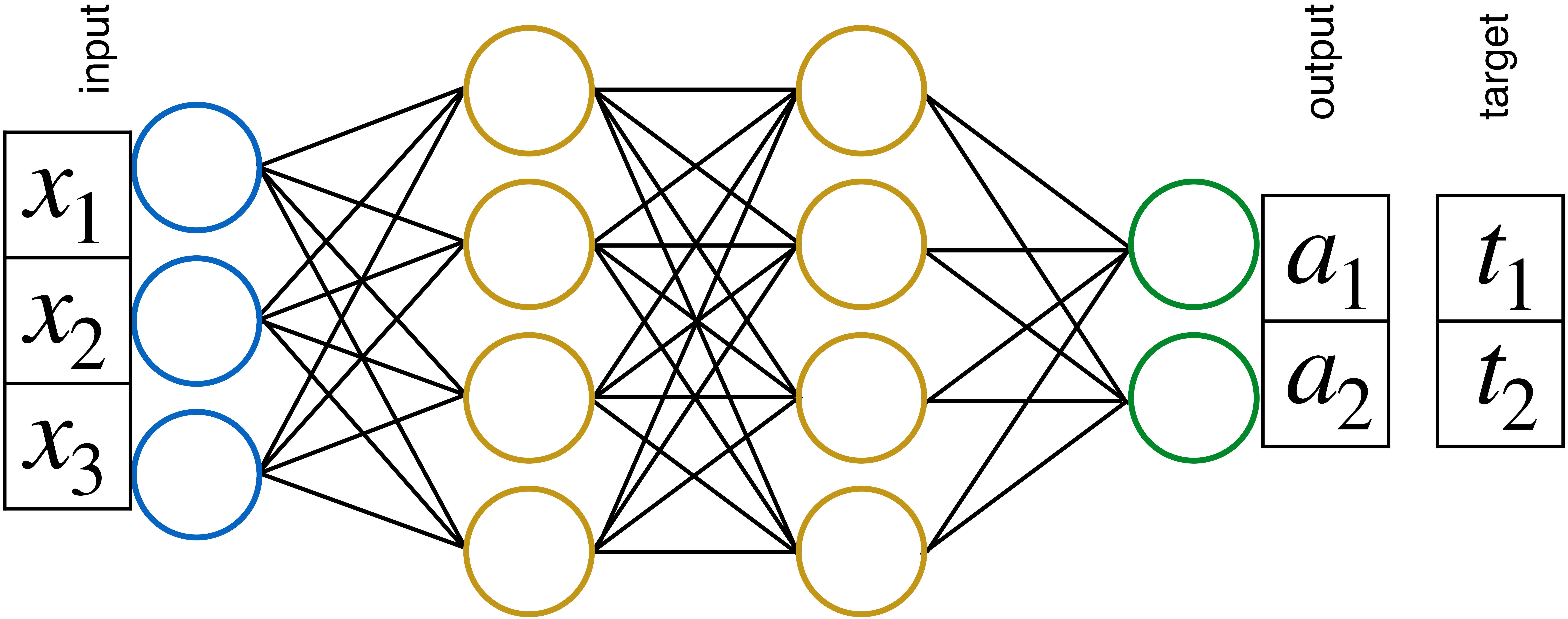
Training a neural network by back-propagation of the error ('backprop')



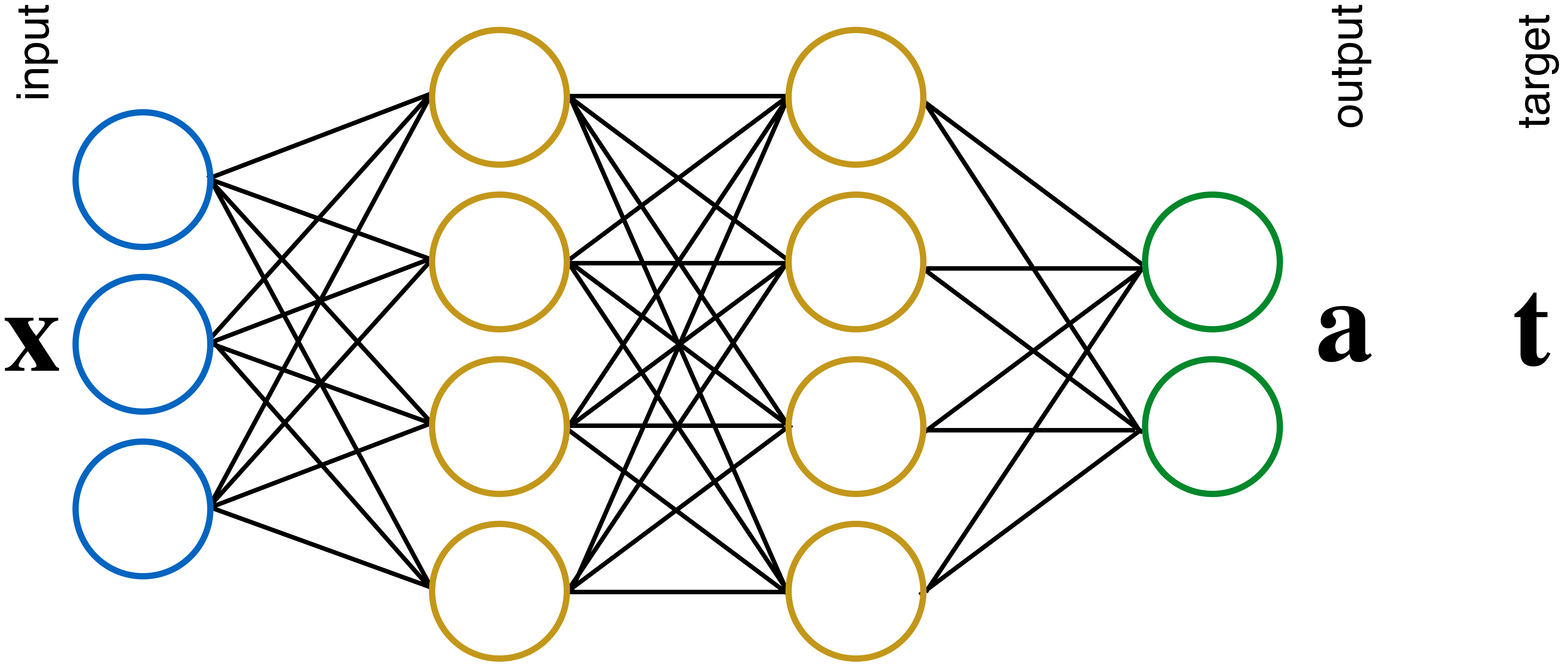
Training a neural network using back-propagation of the error

- what is the objective of training?
- <u>notation</u>
- taking the derivative
- deriving back-propagation

input, output, target

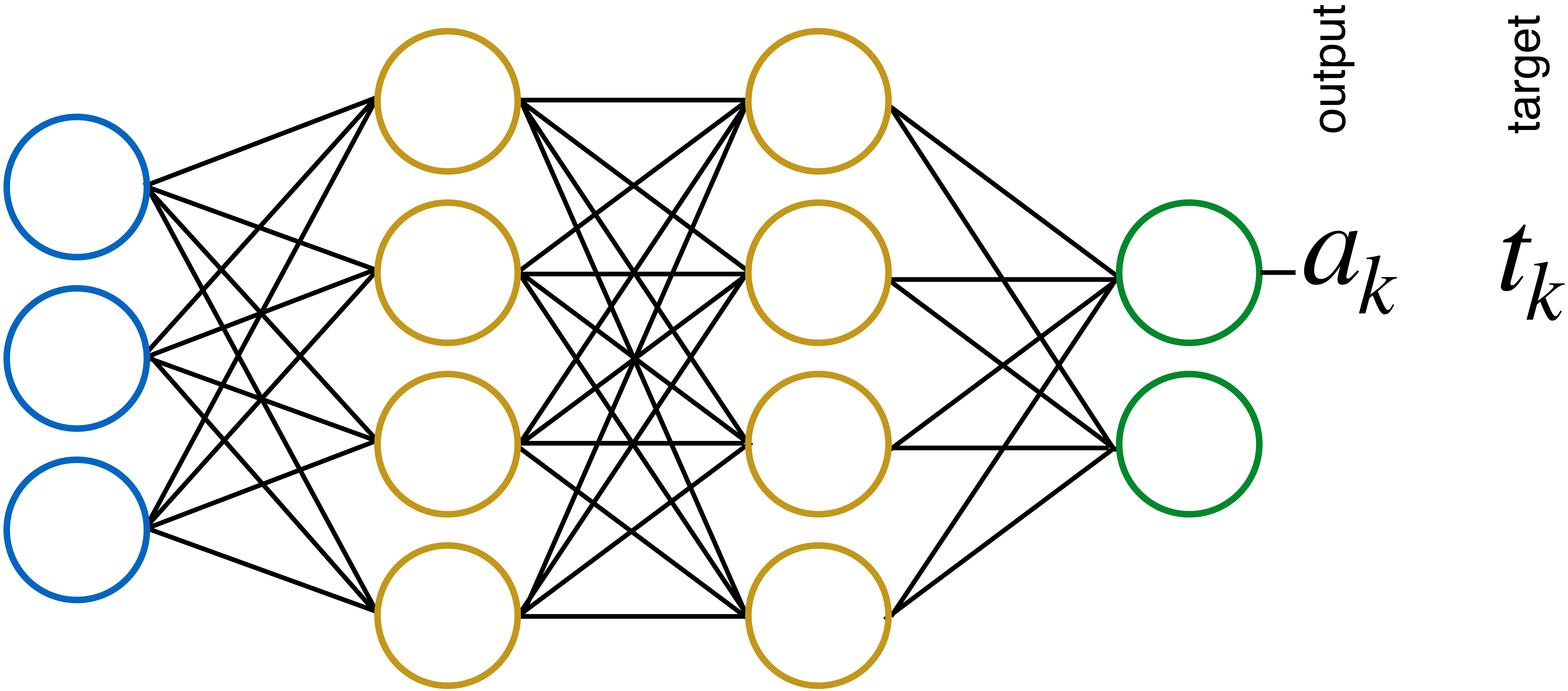


input, output, target - could write as vectors

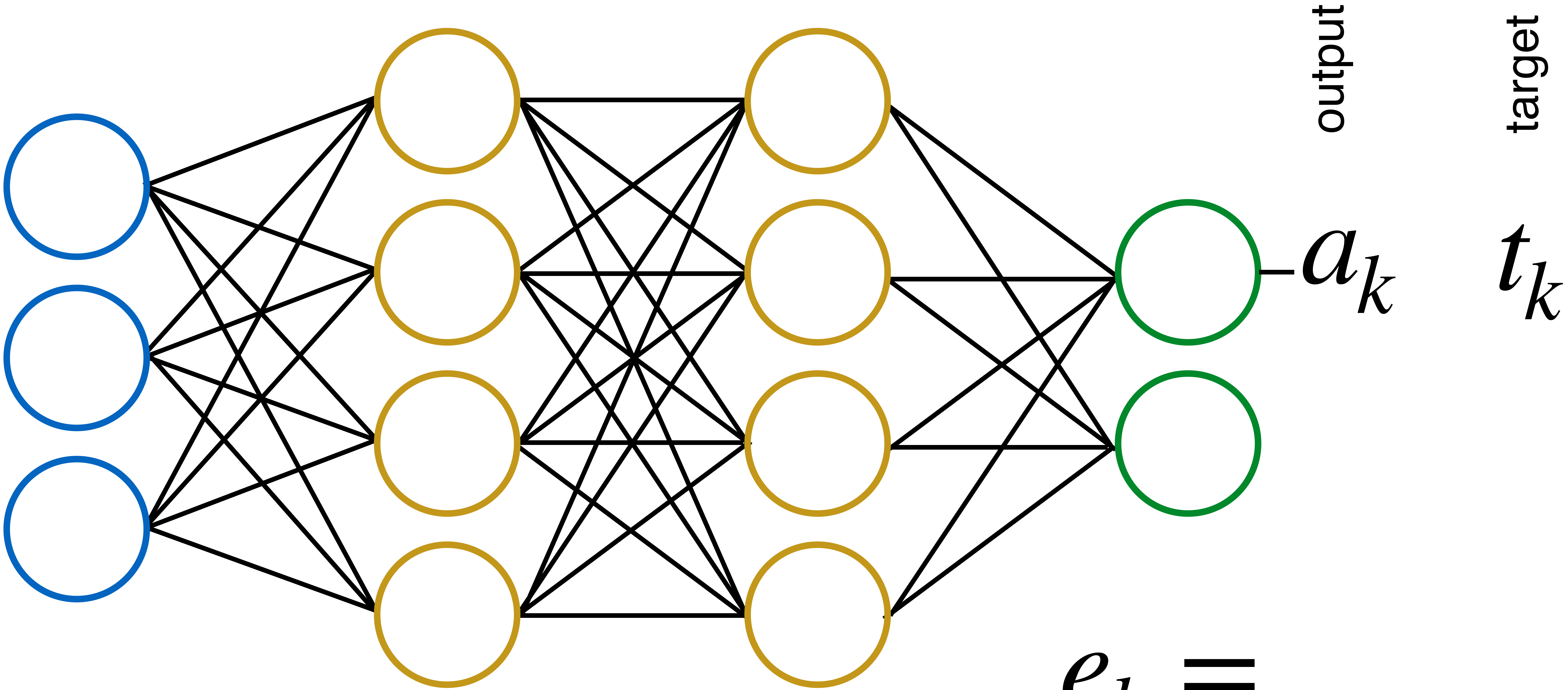


The goal of training is to choose model parameters that minimise error

Each output is the activation of a unit in the output layer



The error at one output



$$e_k =$$

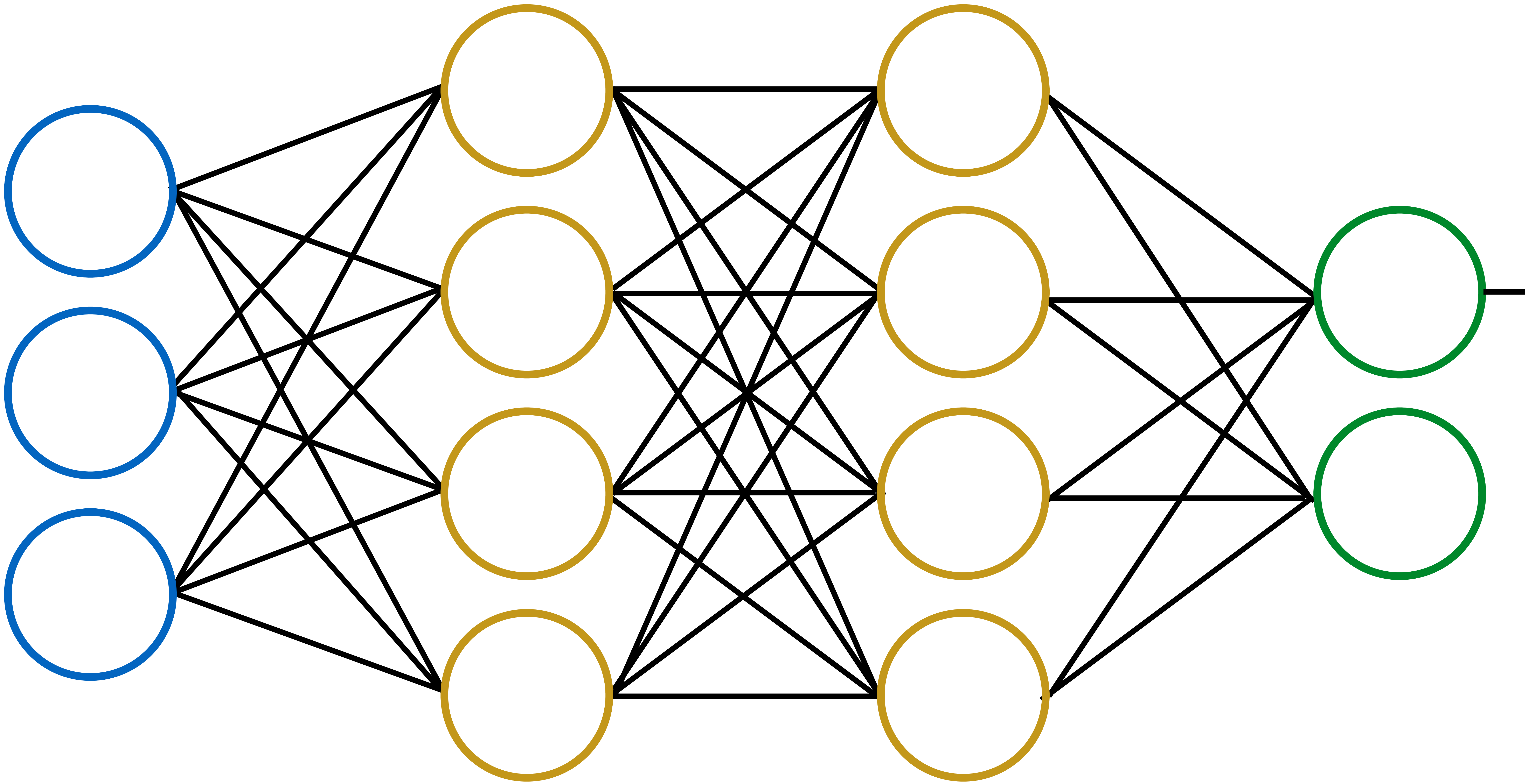
Define the total error to be minimised : E

$$e_k = a_k - t_k$$

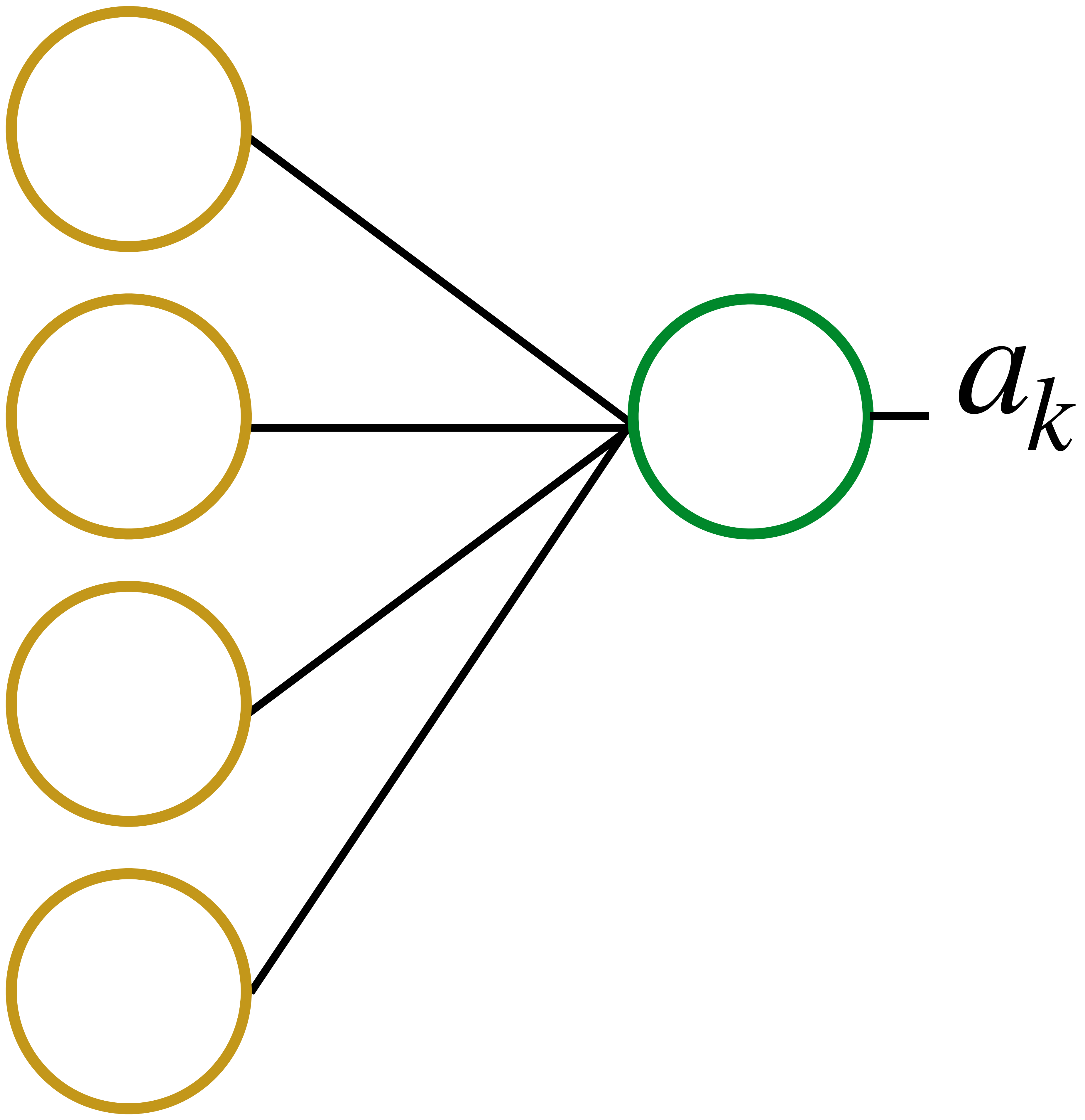
$$E = \sum_{k=1}^{K} (\qquad\qquad)^2$$

$$E = \sum_{k=1}^{K} (\qquad\qquad)^2$$

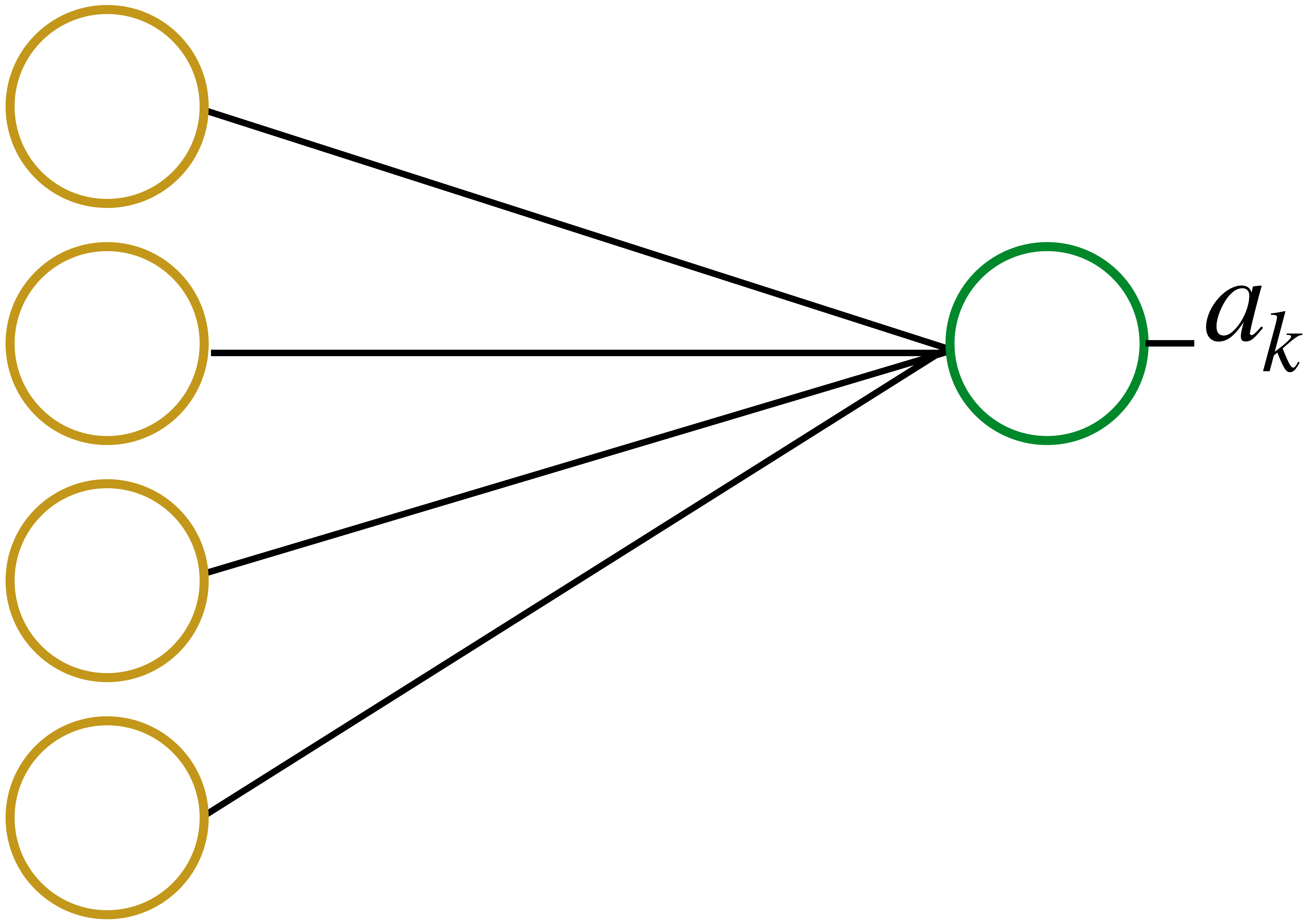
Notation



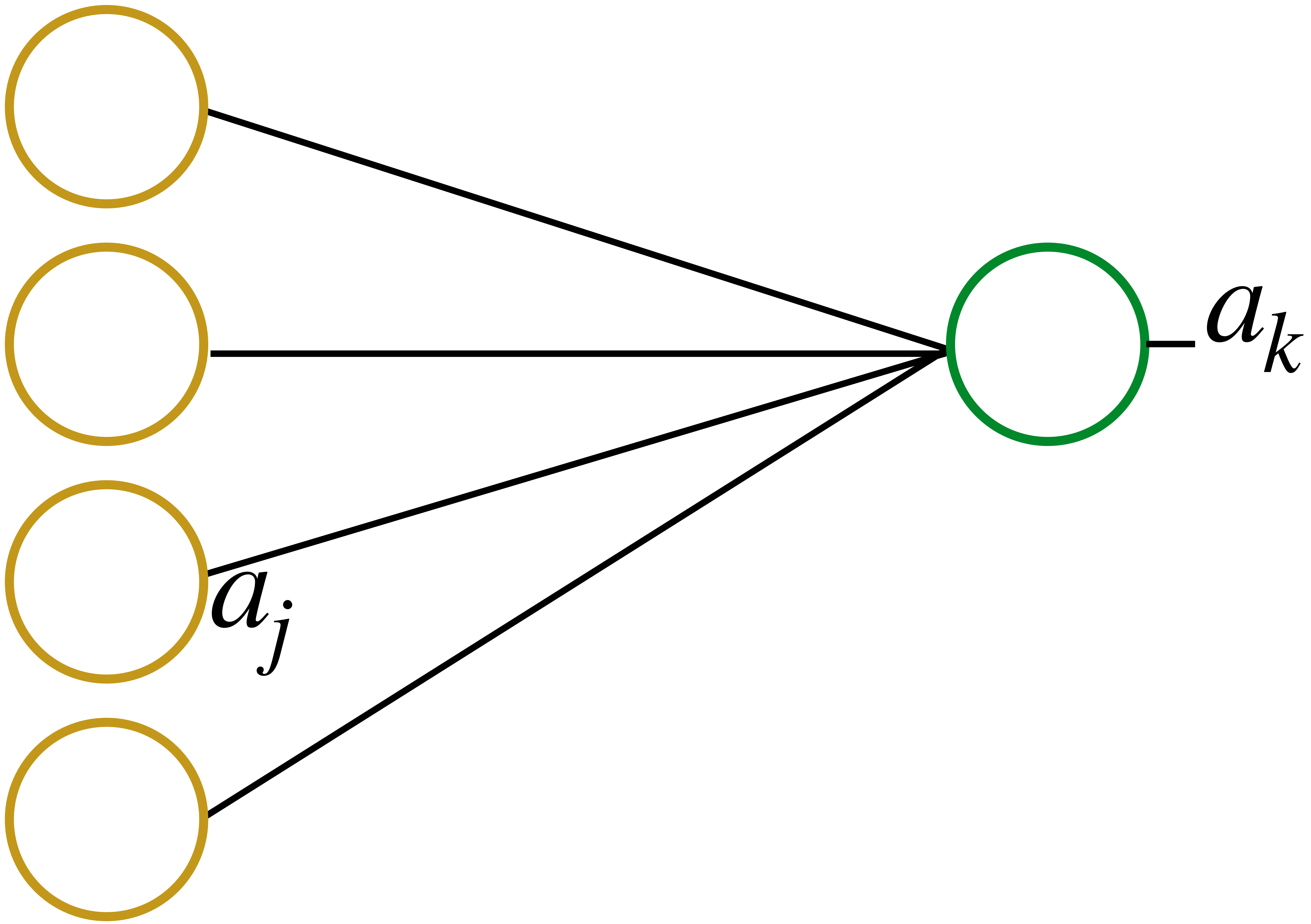
Notation



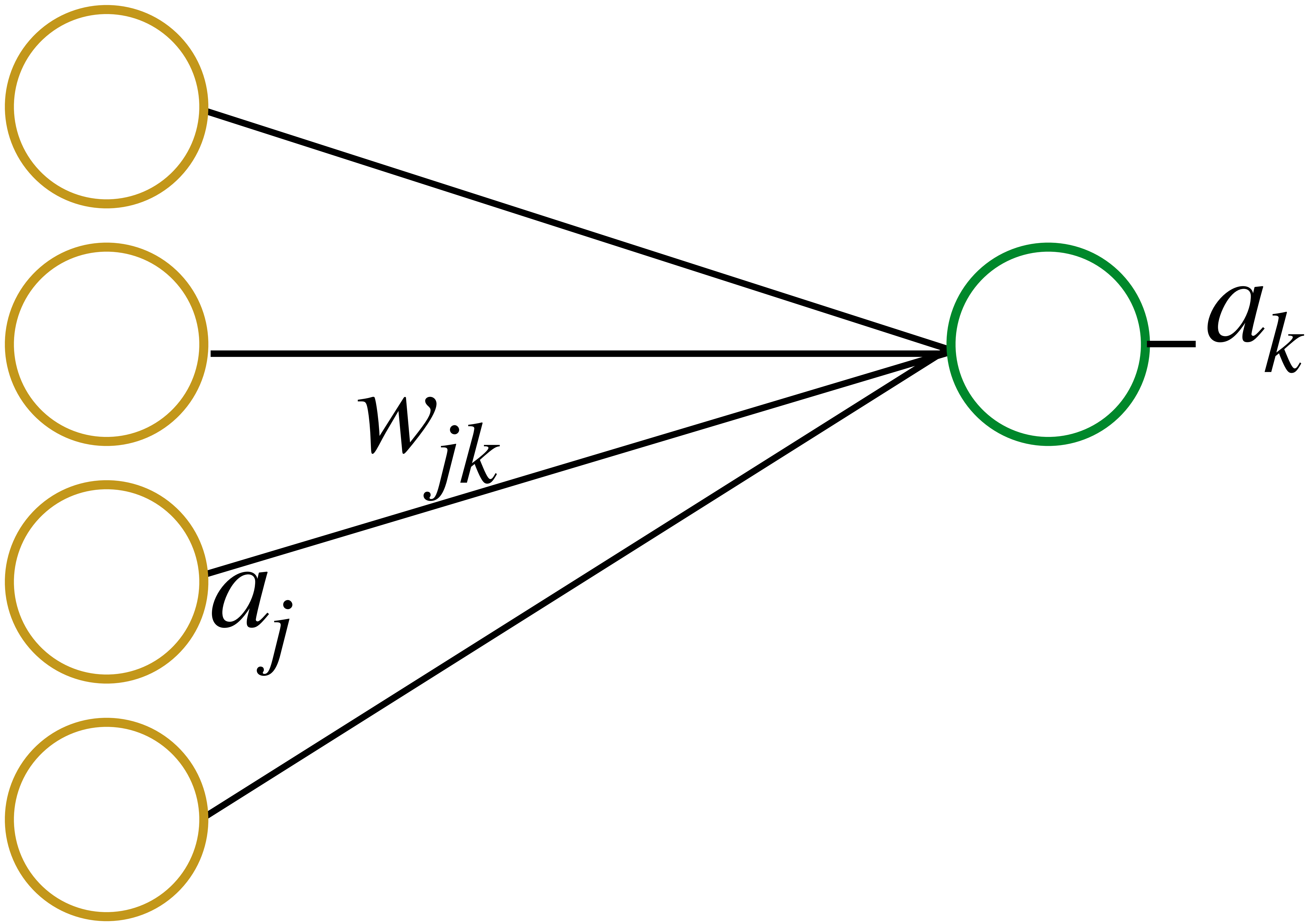
Notation



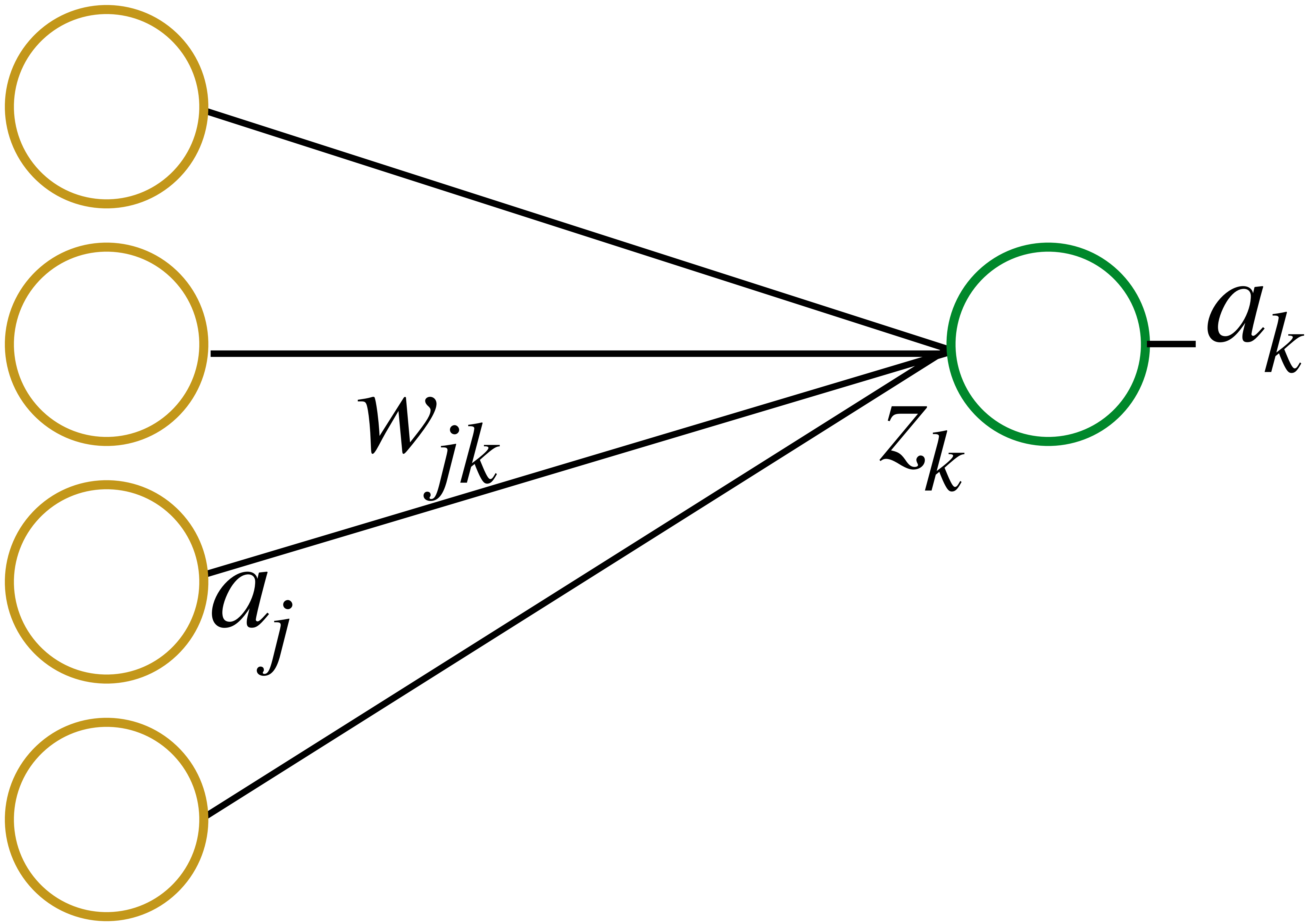
Notation



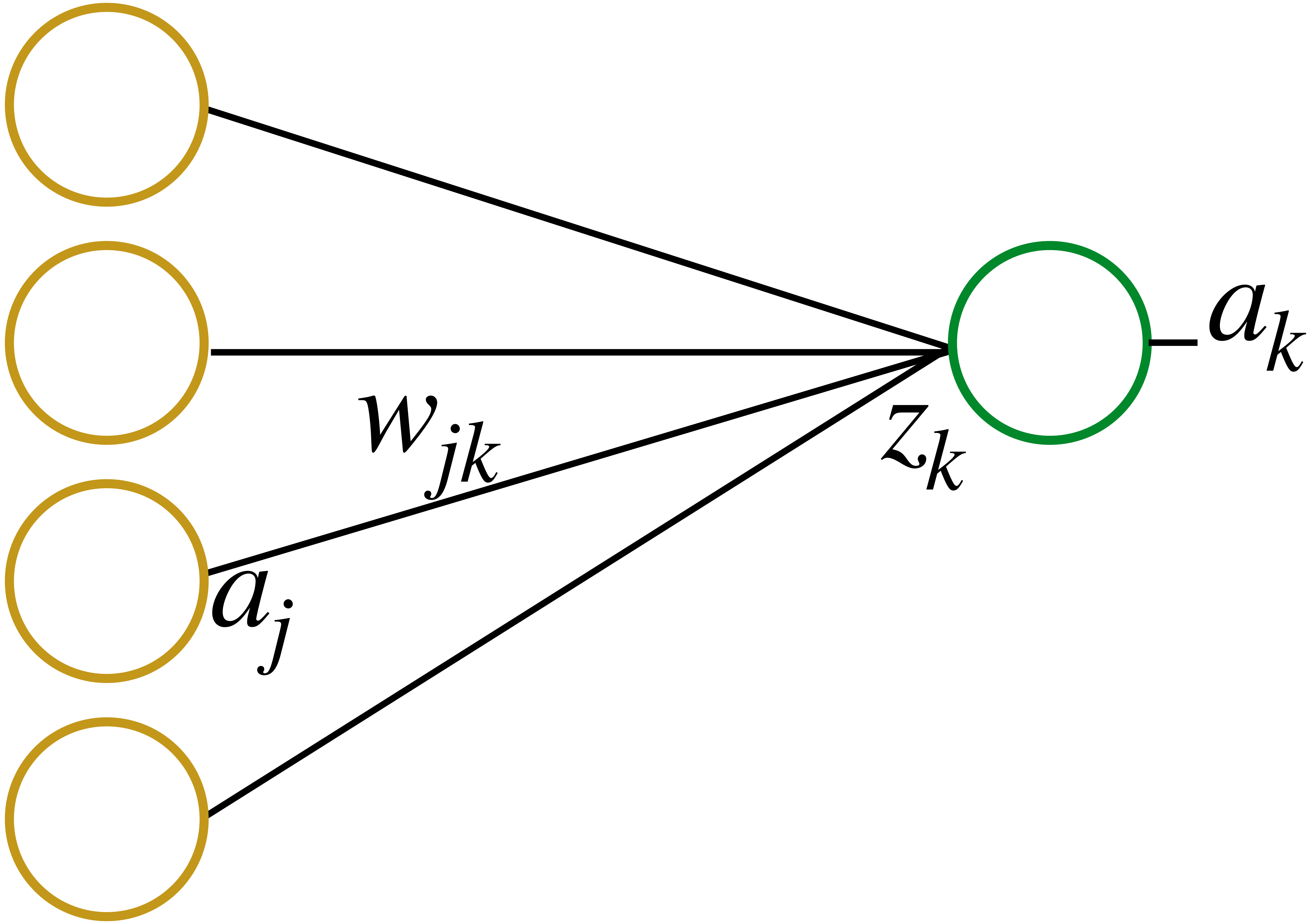
Notation



Notation

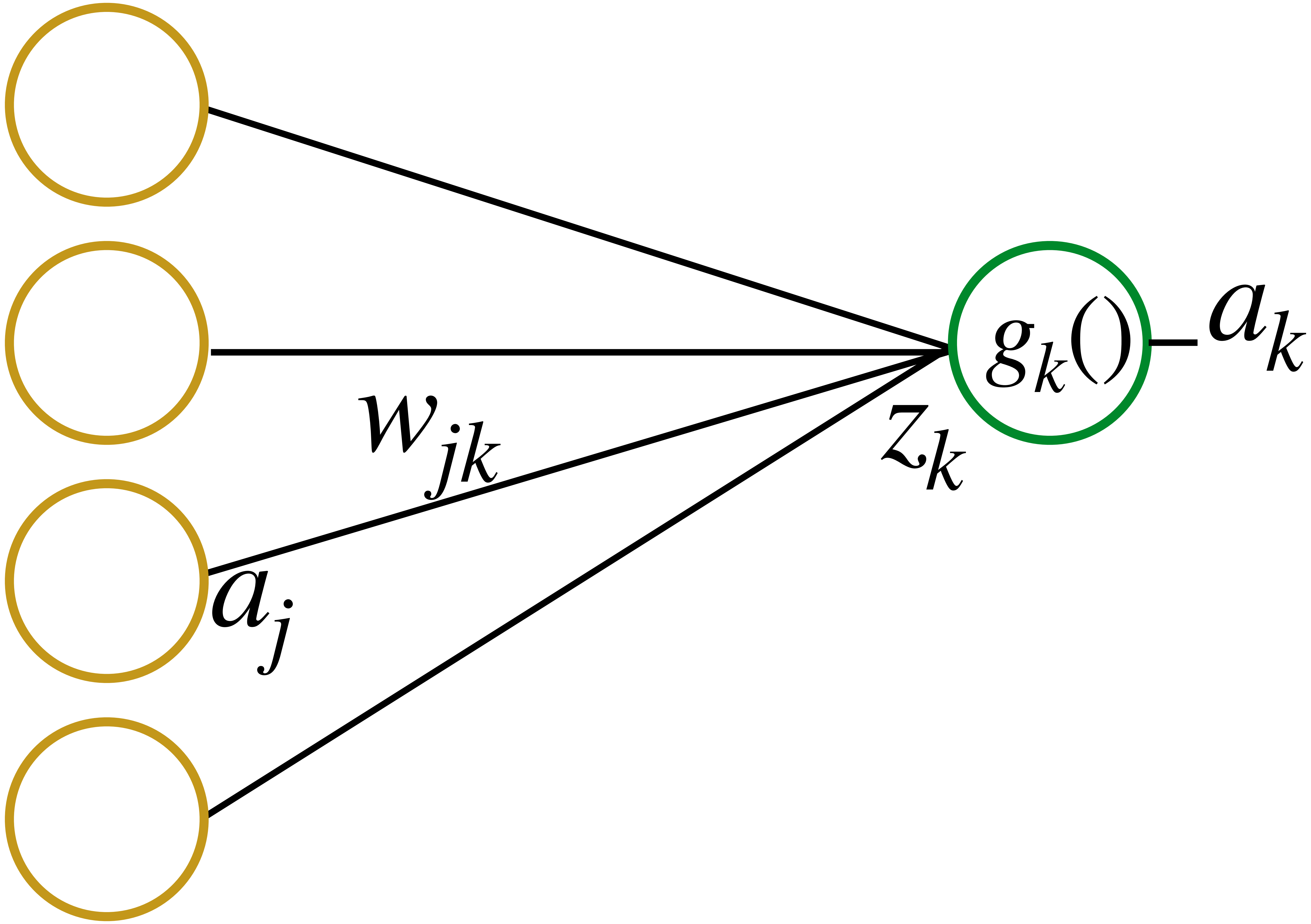


Notation



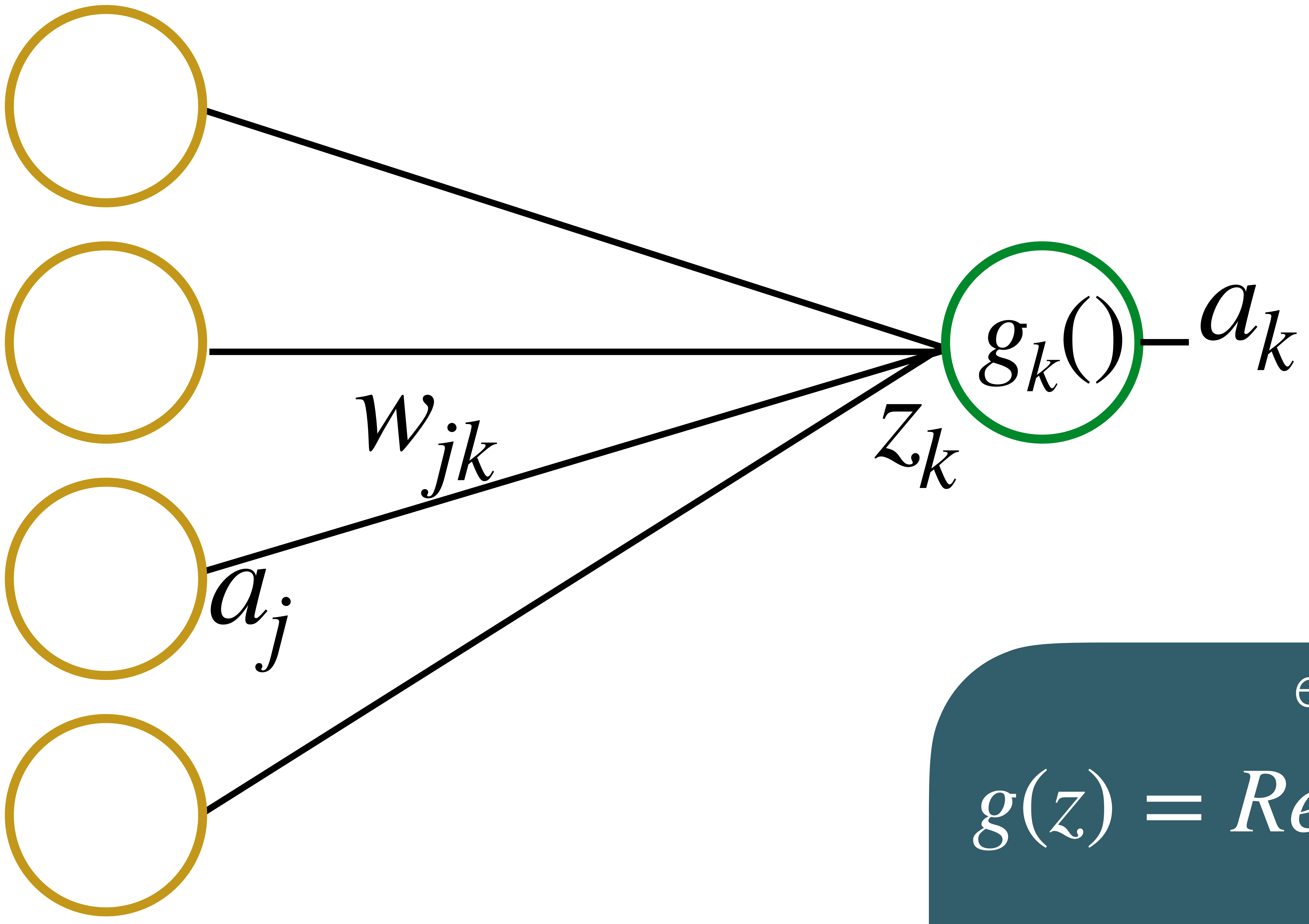
$$z_k = \sum_j a_j \, w_{jk}$$

Notation



$$z_k = \sum_j a_j \, w_{jk}$$

Notation



$$z_k = \sum_j a_j \, w_{jk}$$

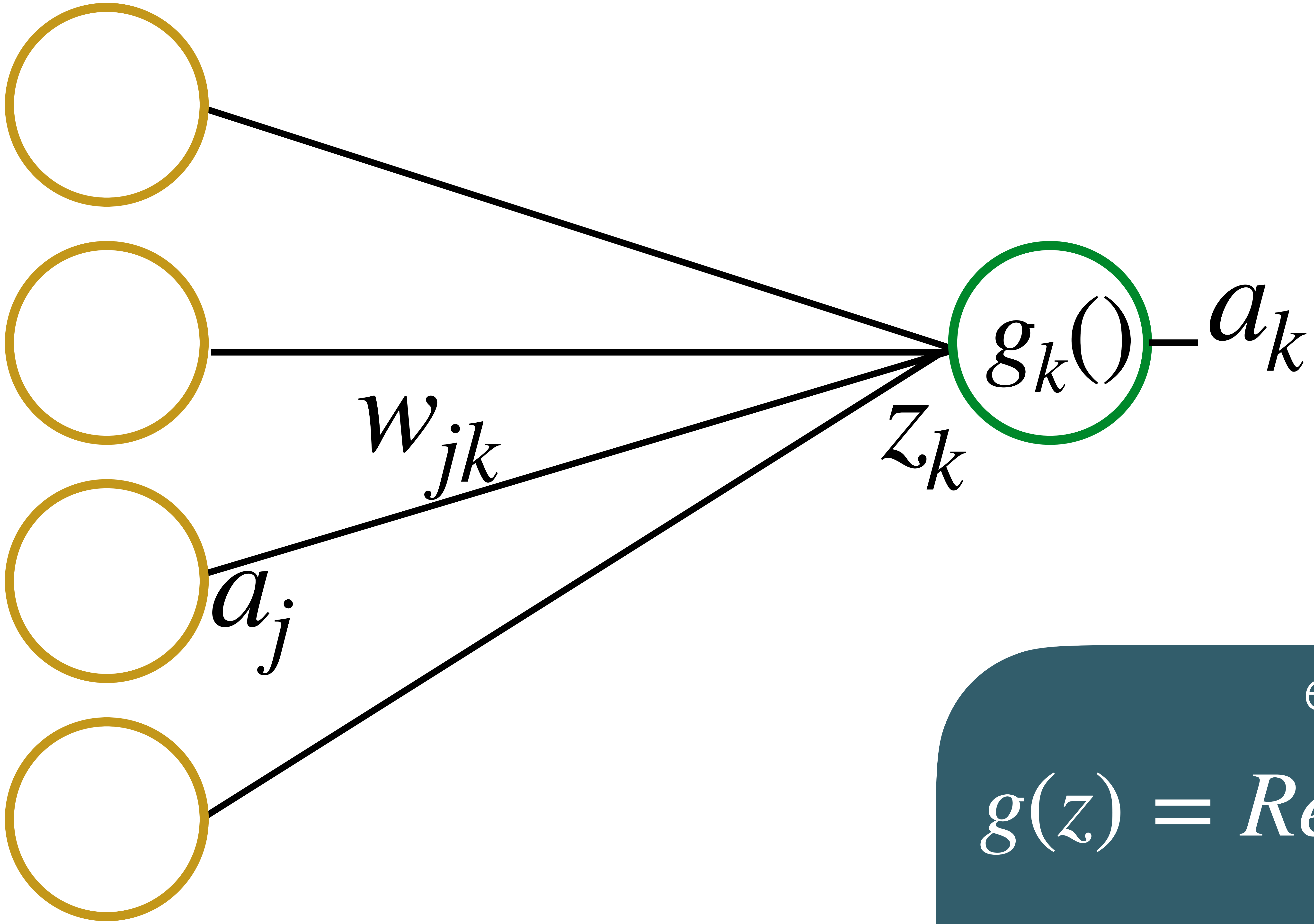
example activation functions

$$g(z) = ReLU(z)$$

$$g(z) = \tanh(z)$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

Notation



$$z_k = \sum_j a_j \, w_{jk}$$

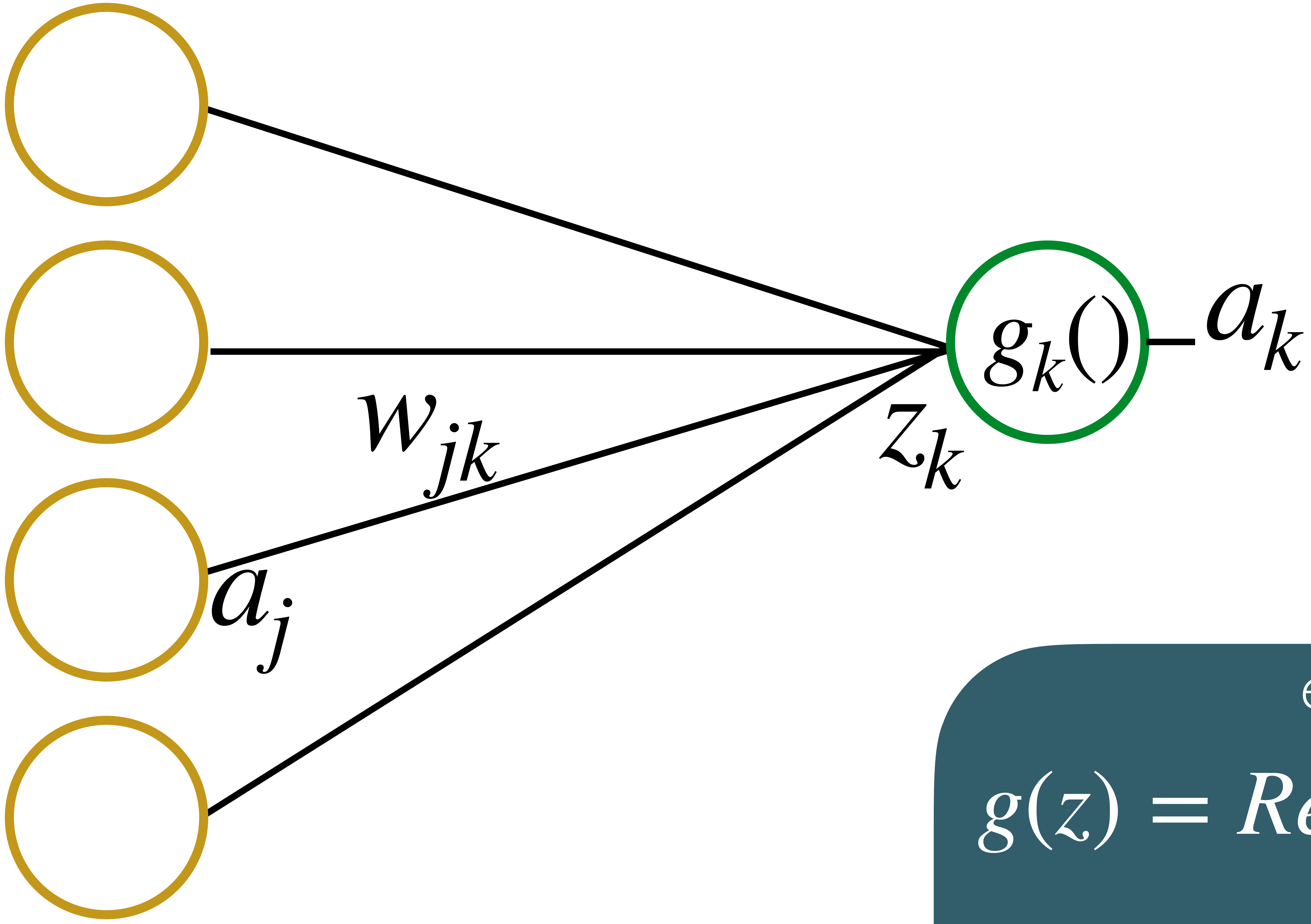
example activation functions

$$g(z) = ReLU(z)$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g(z) = \tanh(z)$$

Notation



$$z_k = \sum_j a_j \, w_{jk}$$

$$a_k = g_k(z_k)$$

example activation functions

$$g(z) = ReLU(z)$$

$$g(z) = \frac{1}{1 + e^{-z}}$$

$$g(z) = \tanh(z)$$

Training a neural network using back-propagation of the error

- what is the objective of training?
- notation
- <u>taking the derivative</u>
- deriving back-propagation

Partial derivative
or, how much does a function change when one variable changes?

$$y = 3a^2 - 4b^3 - 2ac + 8a$$

$$\frac{\partial y}{\partial a} =$$

$$\frac{\partial y}{\partial b} =$$

Partial derivative

or, how much does a function change when one variable changes?

$$y = 3a^2 - 4b^3 - 2ac + 8a$$

$$\frac{\partial y}{\partial a} = 6a \quad - \quad 0 \quad - 2c \quad + 8$$

$$\frac{\partial y}{\partial b} =$$

Partial derivative
or, how much does a function change when one variable changes?

$$y = 3a^2 - 4b^3 - 2ac + 8a$$

$$\frac{\partial y}{\partial a} = 6a \quad - \quad 0 \quad - 2c \quad + 8$$

$$\frac{\partial y}{\partial b} = \qquad - 12b^2$$

Differentiating a sum

$$Y = \sum_{k=1}^{K} m_k \, n_k$$

$$Y = m_1 \, n_1 \; + \; m_2 \, n_2 \; + \; m_3 \, n_3 \; + \; \ldots \; + \; m_K \, n_K$$

$$\frac{\partial Y}{\partial m_3} =$$

Differentiating a sum

$$Y = \sum_{k=1}^{K} m_k \, n_k$$

$$Y = m_1 \, n_1 \; + \; m_2 \, n_2 \; + \; m_3 \, n_3 \; + \; \dots \; + \; m_K \, n_K$$

$$\frac{\partial Y}{\partial m_k} =$$

Any term not involving the variable is constant …. and therefore is zero in the partial derivative

$$Y = \sum_{k=1}^{K} m_k \, n_k$$

$$Y = \sum_{k=1}^{K} \left(m_k^3 \, n_k^2 + n_k^4 \right)$$

$$\frac{\partial Y}{\partial m_k} =$$

$$\frac{\partial Y}{\partial m_k} =$$

Any term not involving the variable is constant …. and therefore is zero in the partial derivative

$$Y = \sum_{k=1}^{K} m_k \, n_k$$

$$\frac{\partial Y}{\partial m_k} = n_k$$

$$Y = \sum_{k=1}^{K} \left(m_k^3 \, n_k^2 + n_k^4 \right)$$

$$\frac{\partial Y}{\partial m_k} =$$

Any term not involving the variable is constant …. and therefore is zero in the partial derivative

$$Y = \sum_{k=1}^{K} m_k\, n_k$$

$$Y = \sum_{k=1}^{K} \left(m_k^3\, n_k^2 \;+\; n_k^4 \right)$$

$$\frac{\partial Y}{\partial m_k} = n_k$$

$$\frac{\partial Y}{\partial m_k} = 3m_k^2\, n_k^2$$

The chain rule

$$E = f(e)$$

$$\frac{\partial E}{\partial w} =$$

The chain rule

$$E = f(e)$$

$$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial e} \frac{\partial e}{\partial w}$$

The chain rule

$$E = \frac{1}{2} \sum_{k=1}^{K} (e_k)^2$$

$$\frac{\partial E}{\partial w_{jk}} =$$

The chain rule

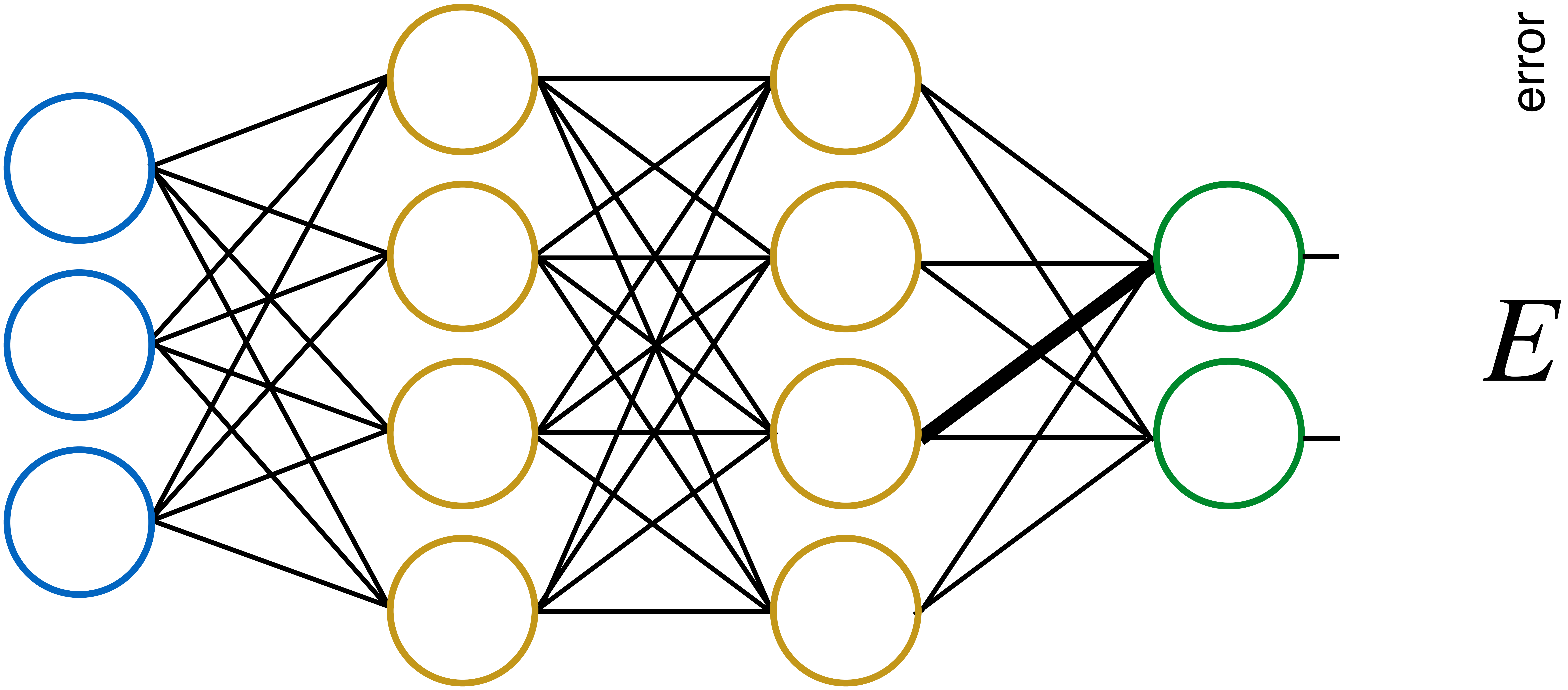
$$E = \frac{1}{2}\sum_{k=1}^{K} (\ e_k\ )^2$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k}\,\frac{\partial e_k}{\partial w_{jk}}$$

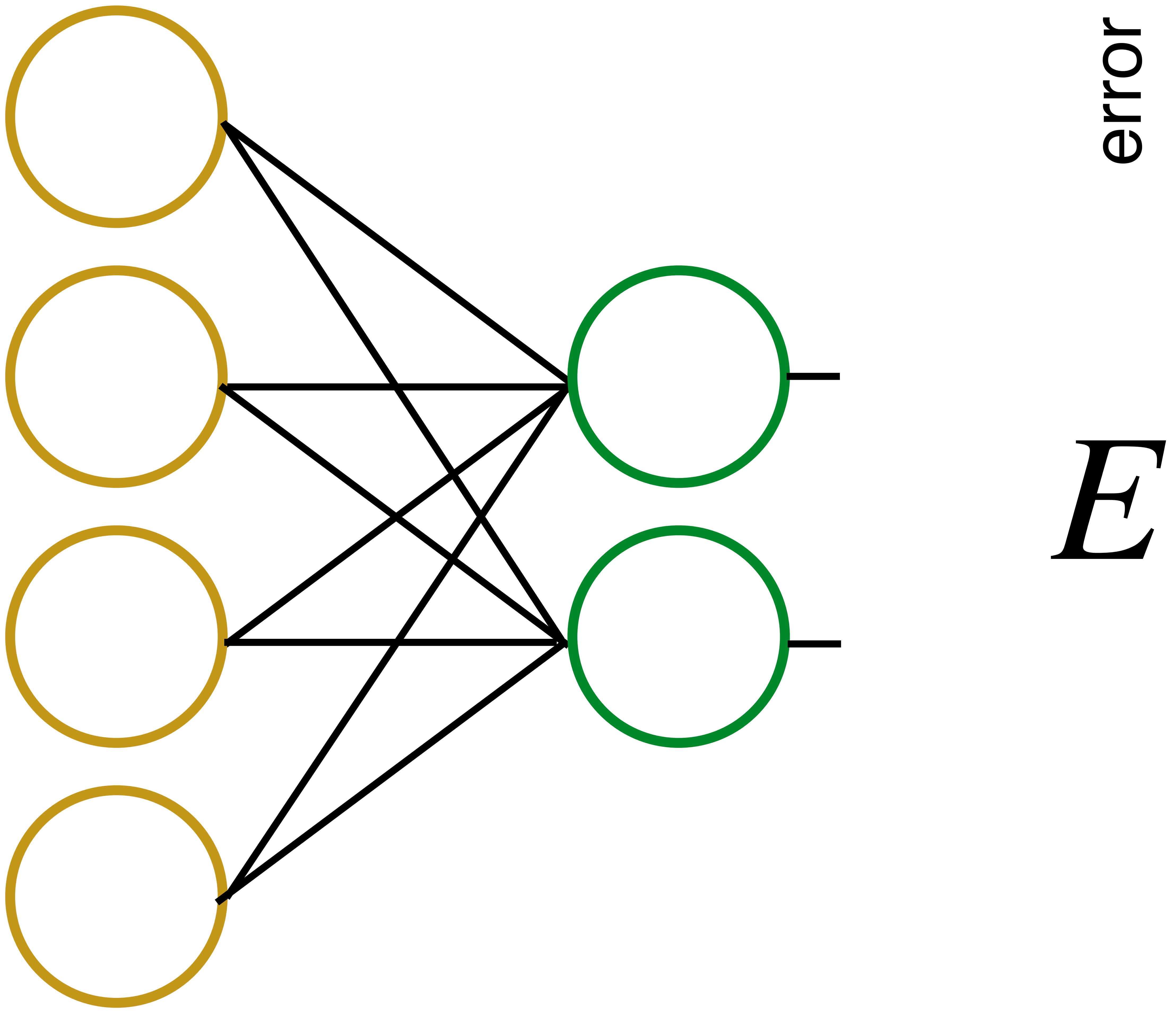
Training a neural network using back-propagation of the error

- what is the objective of training?
- notation
- taking the derivative
- <u>deriving back-propagation</u>

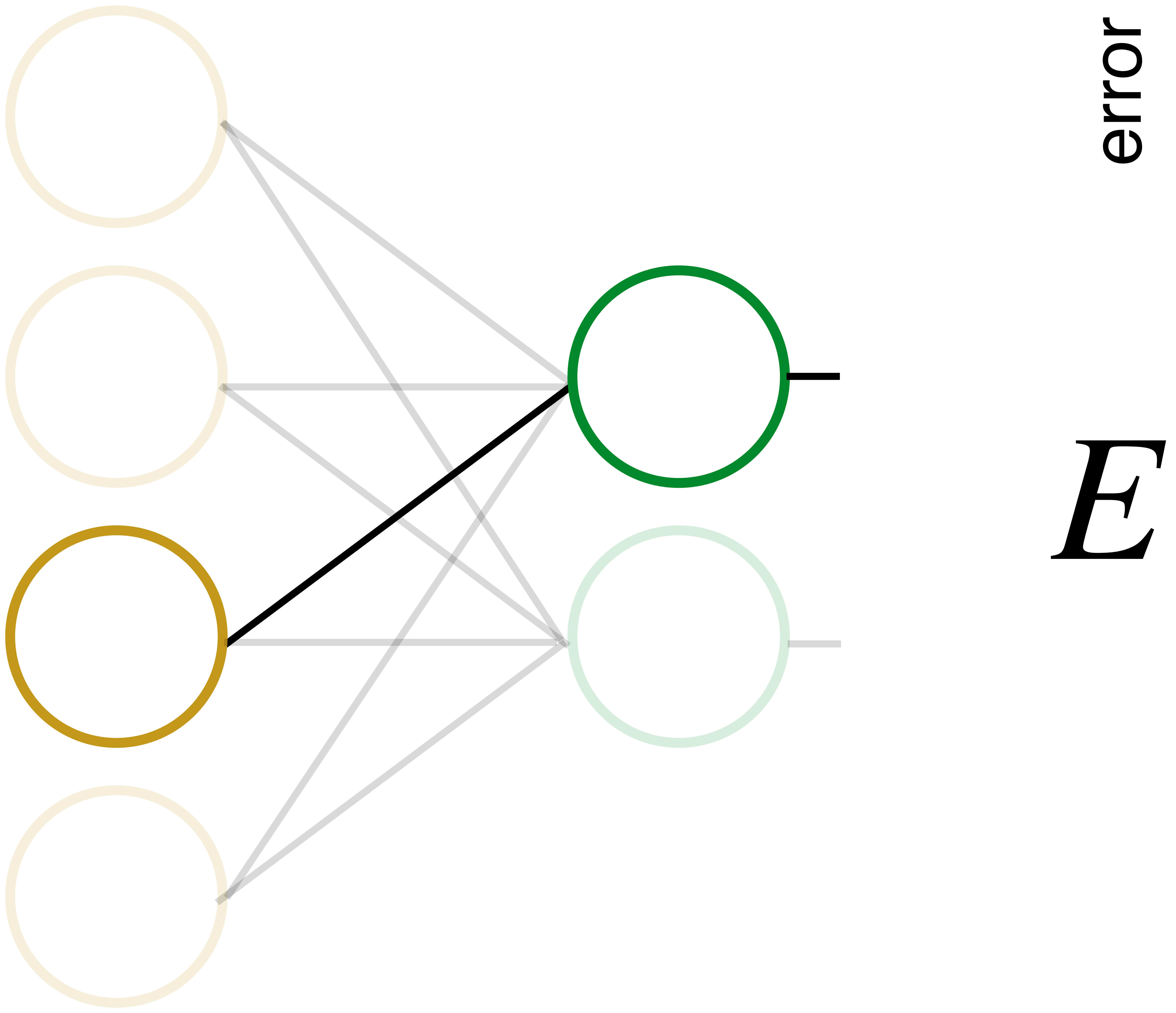
How much would the total error E change, if we changed one weight?



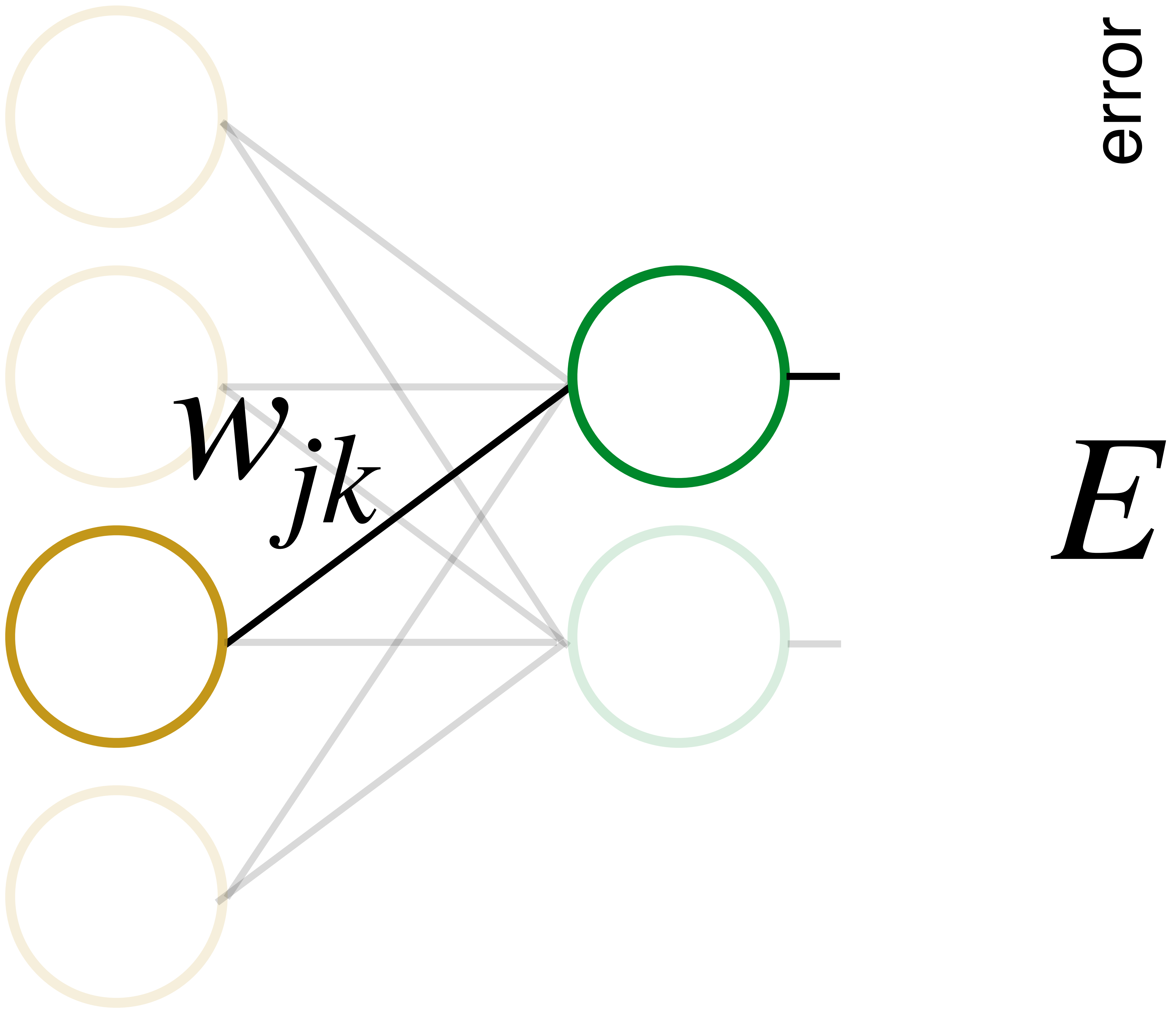
How much would the total error E change, if we changed one weight?



How much would the total error E change, if we changed one weight?

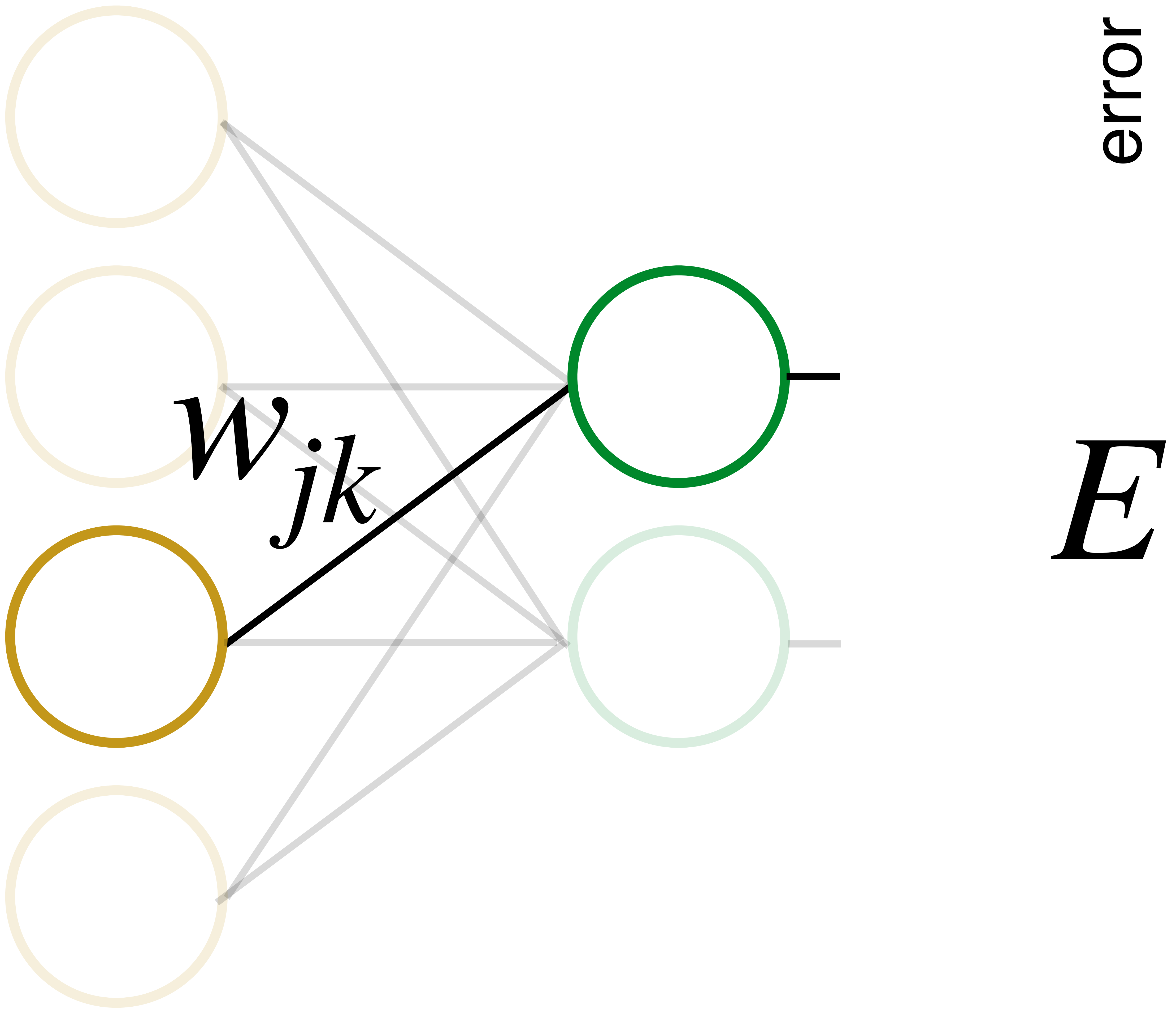


How much would the total error E change, if we changed one weight?



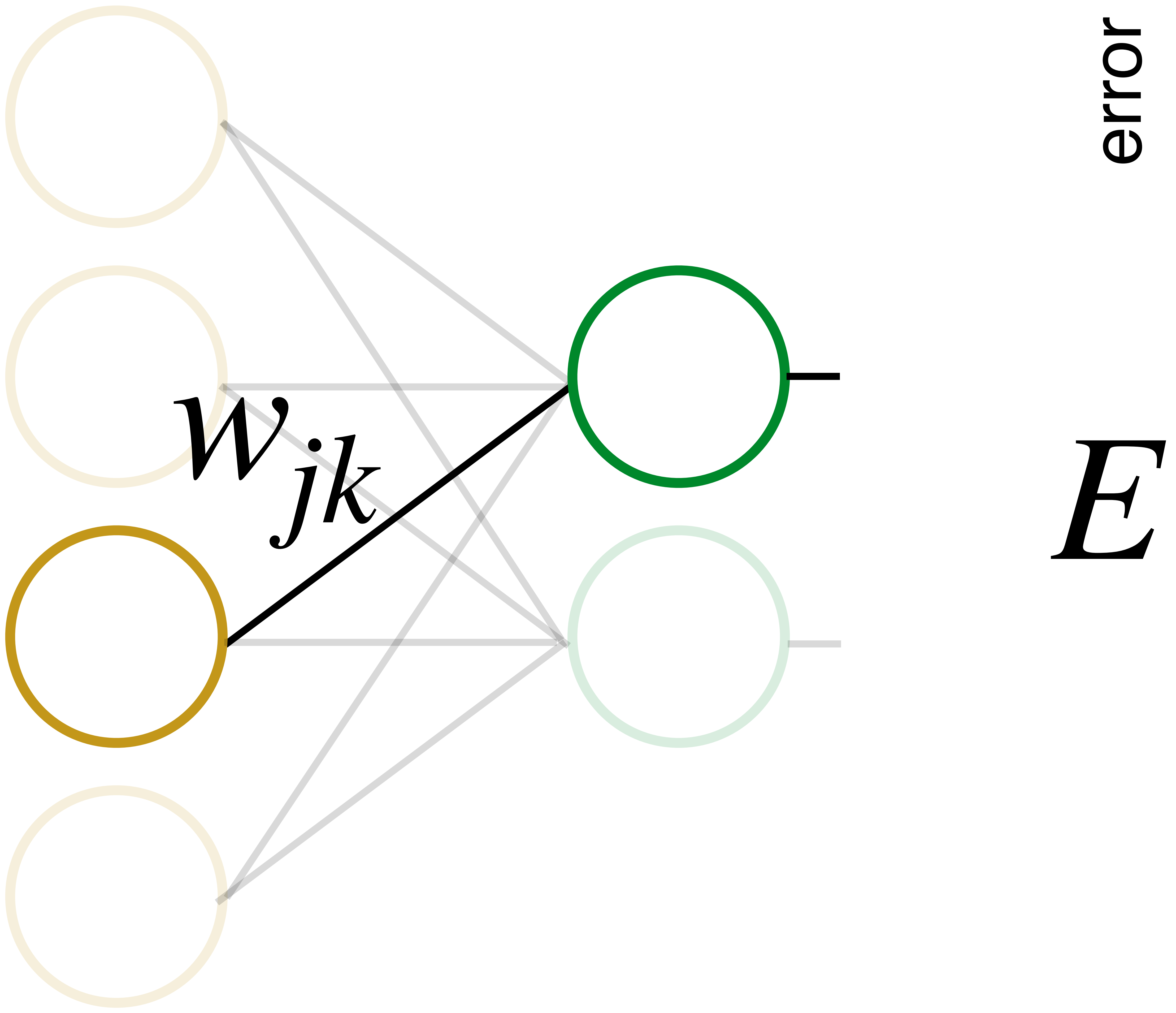
How much would the total error E change, if we changed one weight?

$$\frac{\partial E}{\partial w_{jk}}$$



How much would the total error E change, if we changed one weight?

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial w_{jk}}$$



$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial w_{jk}}$$

$$E = \frac{1}{2} \sum_k (\ e_k\ )^2$$

$$\frac{\partial E}{\partial e_k} =$$

$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k}\frac{\partial e_k}{\partial w_{jk}}$$

$$E = \frac{1}{2}\sum_k (\ e_k\ )^2$$

$$\frac{\partial E}{\partial e_k} = \quad e_k$$

$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial w_{jk}}$$

$$E = \frac{1}{2} \sum_k (e_k)^2$$

$$\frac{\partial E}{\partial e_k} = \quad e_k = a_k - t_k$$

$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial w_{jk}}$$

$$E = \frac{1}{2} \sum_k (e_k)^2$$

$$\frac{\partial E}{\partial e_k} = e_k = a_k - t_k$$

$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} = \frac{\partial a_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial e_k}{\partial w_{jk}}$$

$$E = \frac{1}{2}\sum_k (e_k)^2$$

$$\frac{\partial E}{\partial e_k} = e_k = a_k - t_k$$

$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} = \frac{\partial a_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

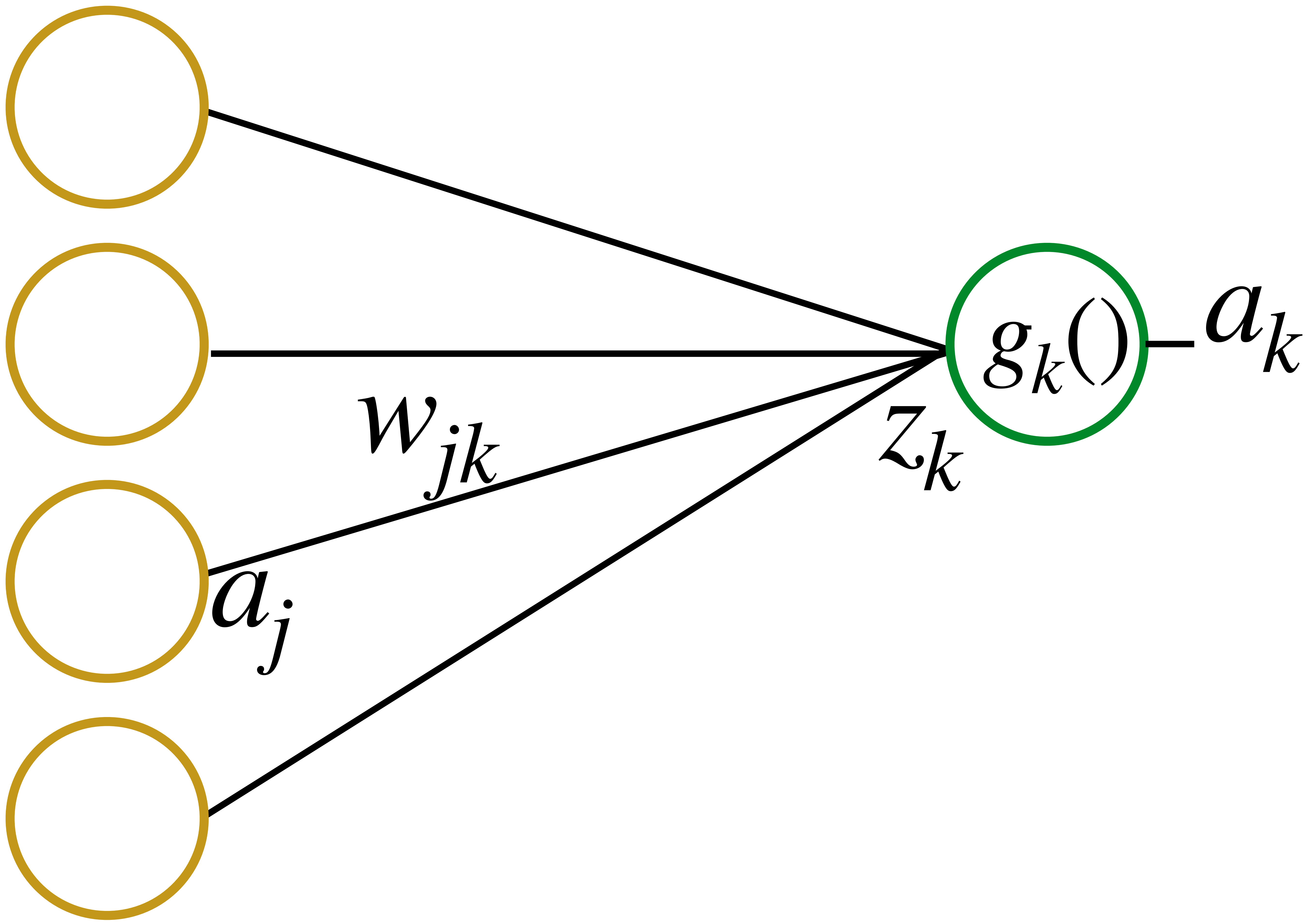
$$E = \frac{1}{2}\sum_k (e_k)^2$$

$$\frac{\partial E}{\partial e_k} = e_k = a_k - t_k$$

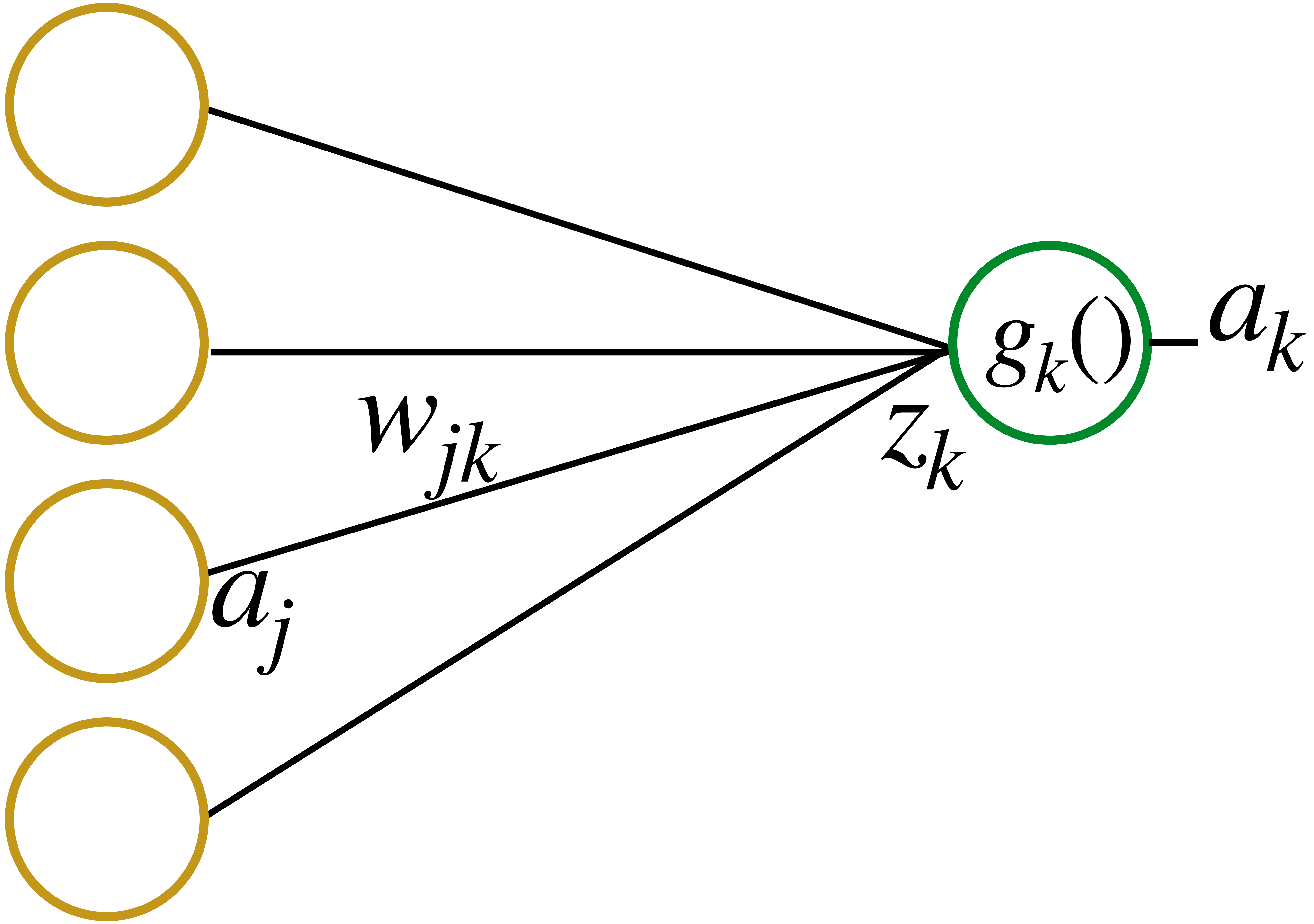
$$e_k = a_k - t_k$$

$$\frac{\partial e_k}{\partial w_{jk}} = \frac{\partial a_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

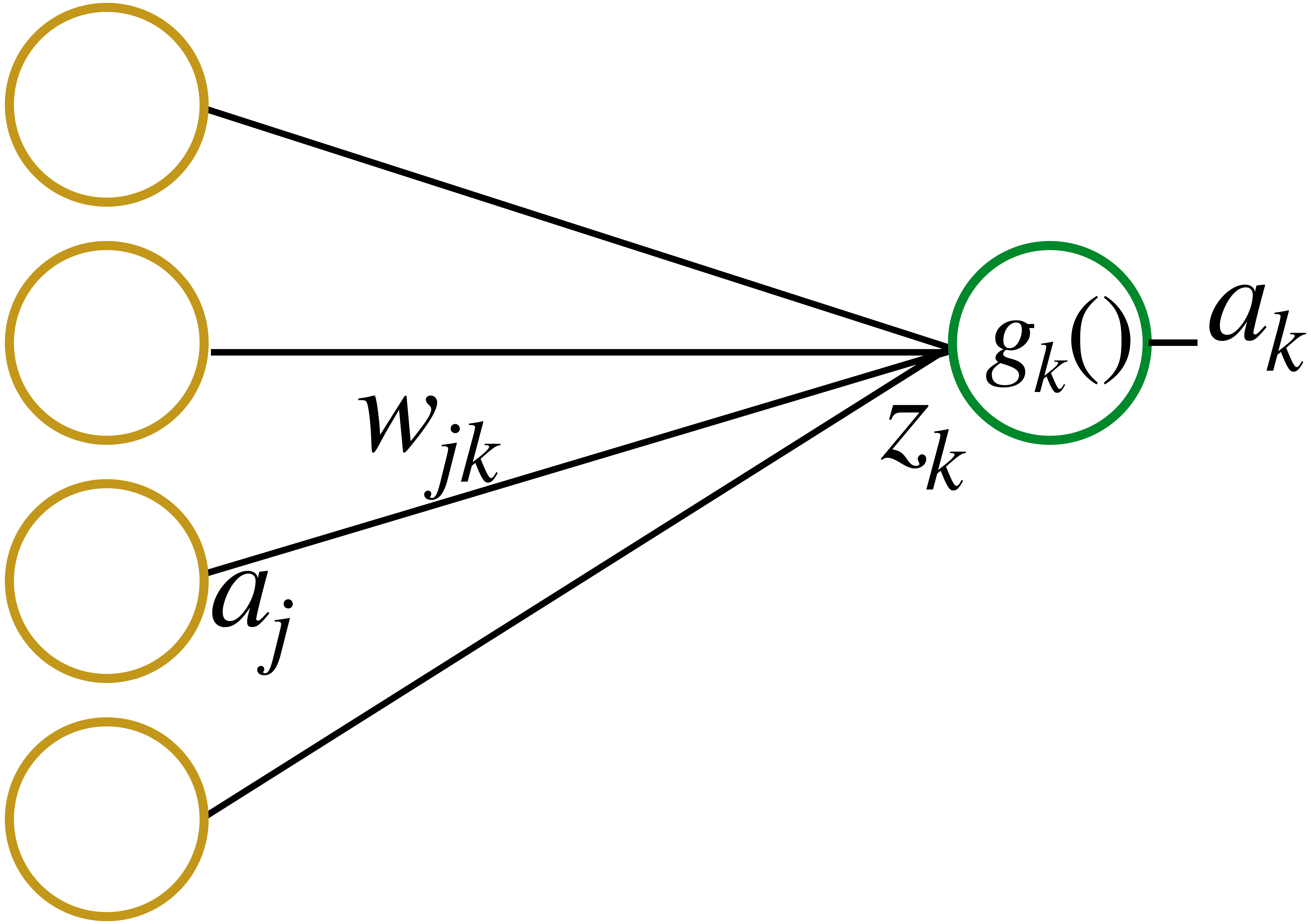


$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$



$$z_k = \sum_j a_j w_{jk}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$



$$a_k = g_k(z_k)$$

$$z_k = \sum_j a_j\, w_{jk}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

$$a_k = g_k(z_k)$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

$$a_k = g_k(z_k)$$

$$\frac{\partial a_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

$$a_k = g_k(z_k)$$

$$\frac{\partial a_k}{\partial w_{jk}} = \frac{\partial g_k(z_k)}{\partial z_k} \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

$$a_k = g_k(z_k)$$

$$\frac{\partial a_k}{\partial w_{jk}} = \frac{\partial g_k(z_k)}{\partial z_k} \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial a_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \frac{\partial a_k}{\partial w_{jk}}$$

$$a_k = g_k(z_k)$$

$$\frac{\partial a_k}{\partial w_{jk}} = \frac{\partial g_k(z_k)}{\partial z_k} \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial a_k}{\partial w_{jk}} = g_k'(z_k) \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$

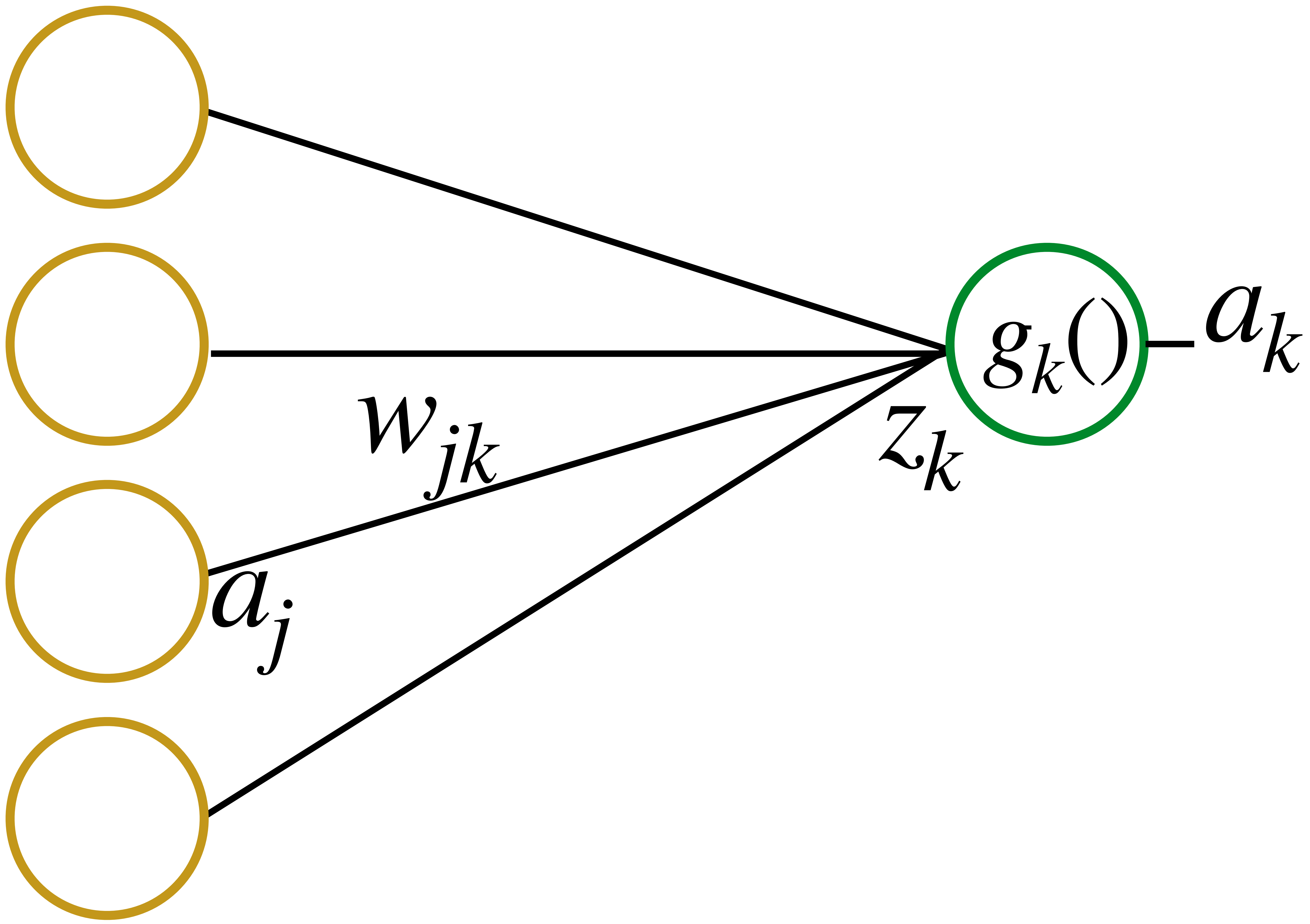
$$a_k = g_k(z_k)$$

$$\frac{\partial a_k}{\partial w_{jk}} = \frac{\partial g_k(z_k)}{\partial z_k} \frac{\partial z_k}{\partial w_{jk}}$$

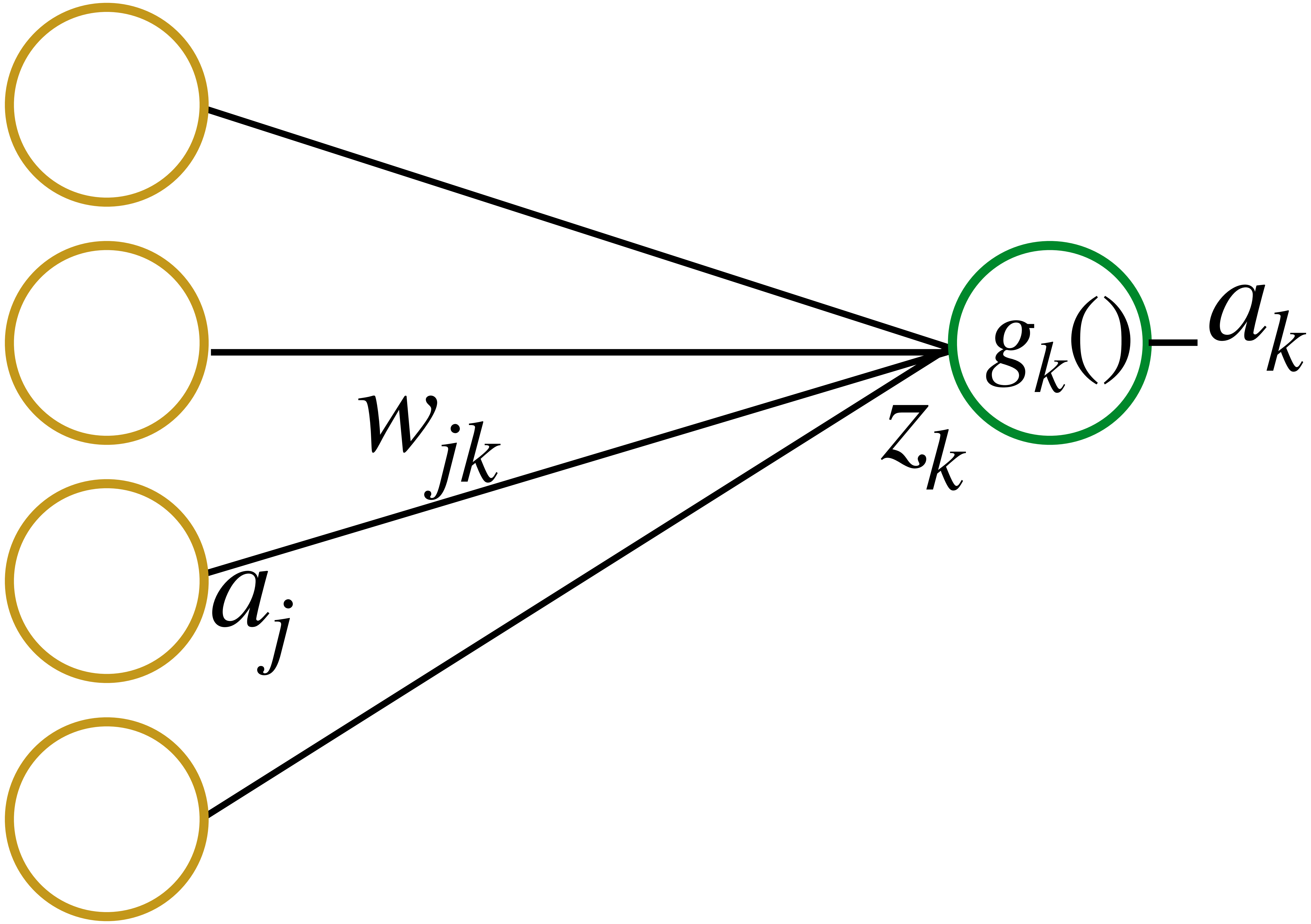
$$\frac{\partial a_k}{\partial w_{jk}} = g_k'(z_k) \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$



$$\frac{\partial E}{\partial w_{jk}} = e_k \; g_k'(z_k) \; \frac{\partial z_k}{\partial w_{jk}}$$



$$z_k = \sum_j a_j \, w_{jk}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$

$$z_k = \sum_j a_j \, w_{jk}$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$

$$z_k = \sum_j a_j \, w_{jk}$$

$$\frac{\partial z_k}{\partial w_{jk}} =$$

$$\frac{\partial E}{\partial w_{jk}} = e_k \quad g_k'(z_k) \quad \frac{\partial z_k}{\partial w_{jk}}$$

$$z_k = \sum_j a_j \, w_{jk}$$

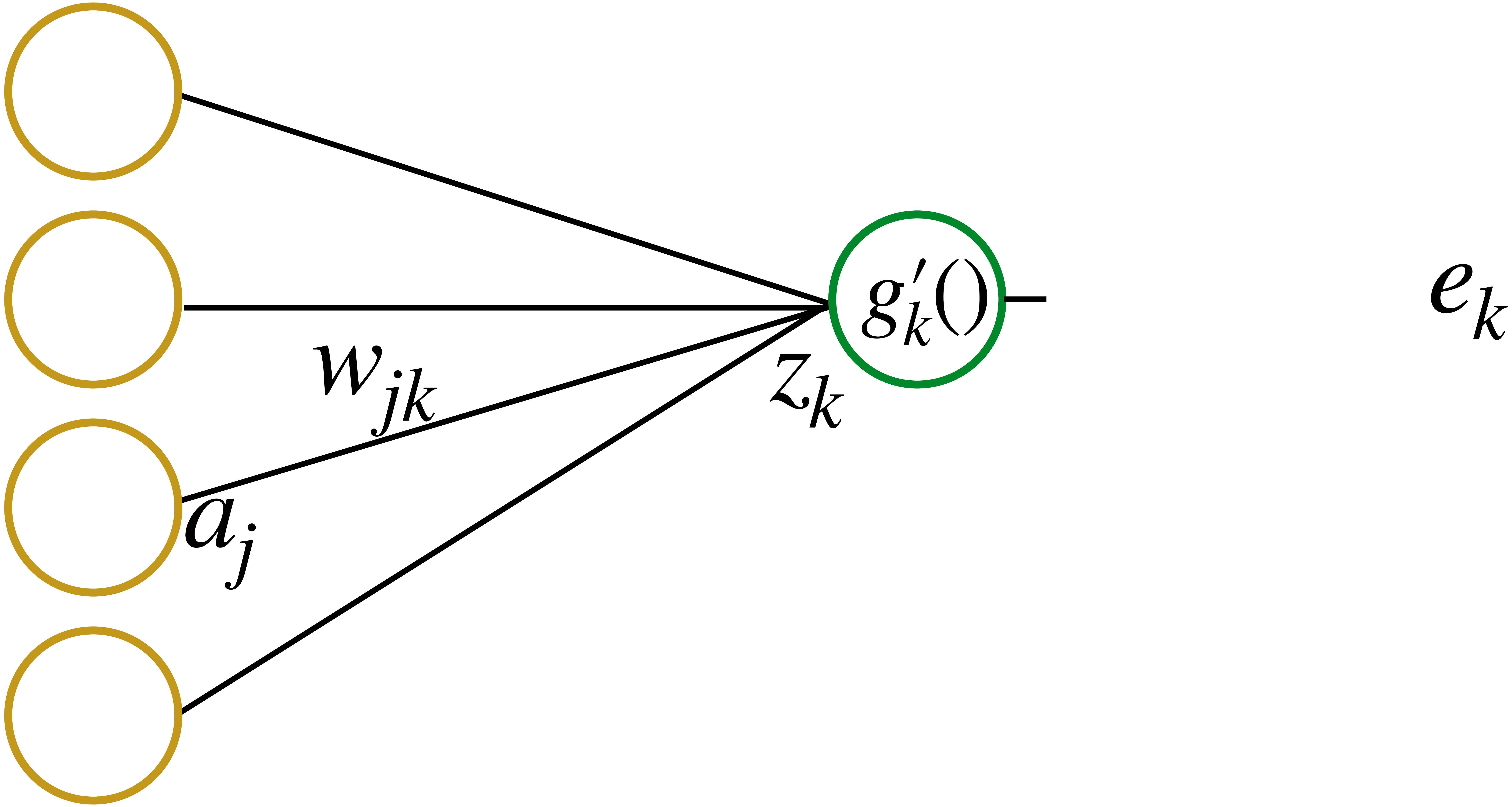
$$\frac{\partial z_k}{\partial w_{jk}} = a_j$$

$$\frac{\partial E}{\partial w_{jk}} = \; e_k \;\; g_k'(z_k) \;\; a_j$$

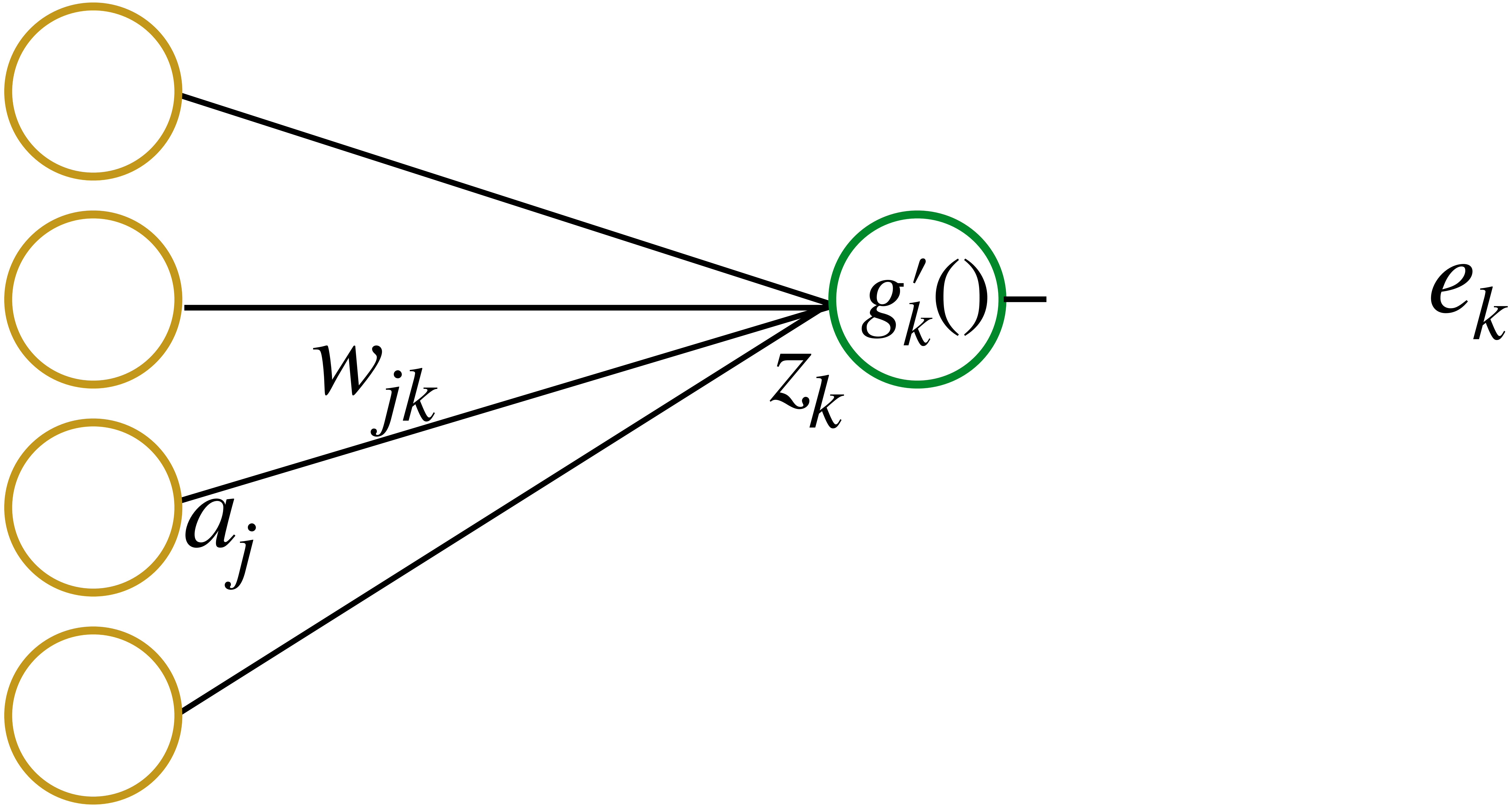
$$z_k = \sum_j a_j \, w_{jk}$$

$$\frac{\partial z_k}{\partial w_{jk}} = a_j$$

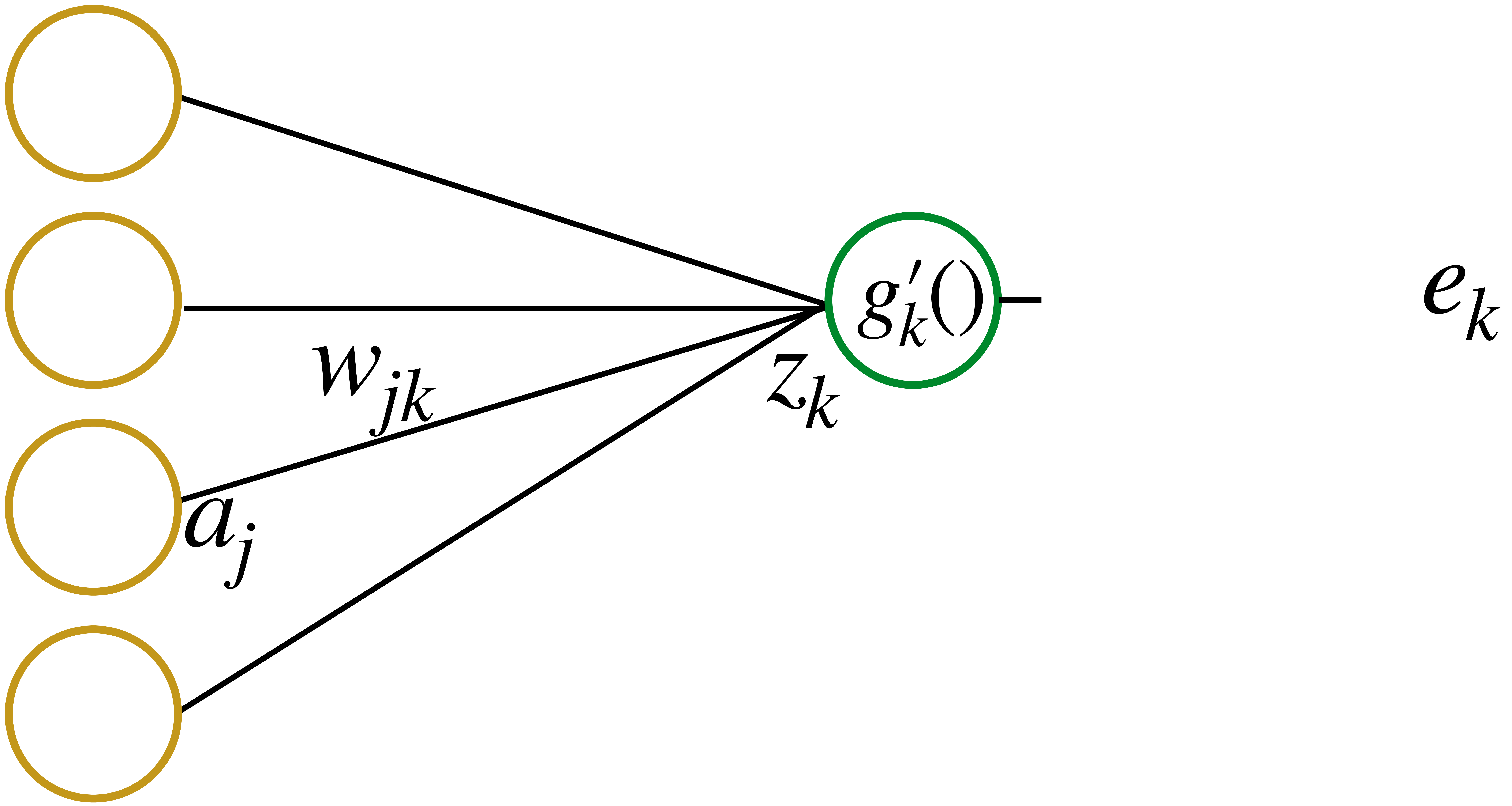
$$\frac{\partial E}{\partial w_{jk}} = e_k \;\; g_k'(z_k) \;\; a_j$$



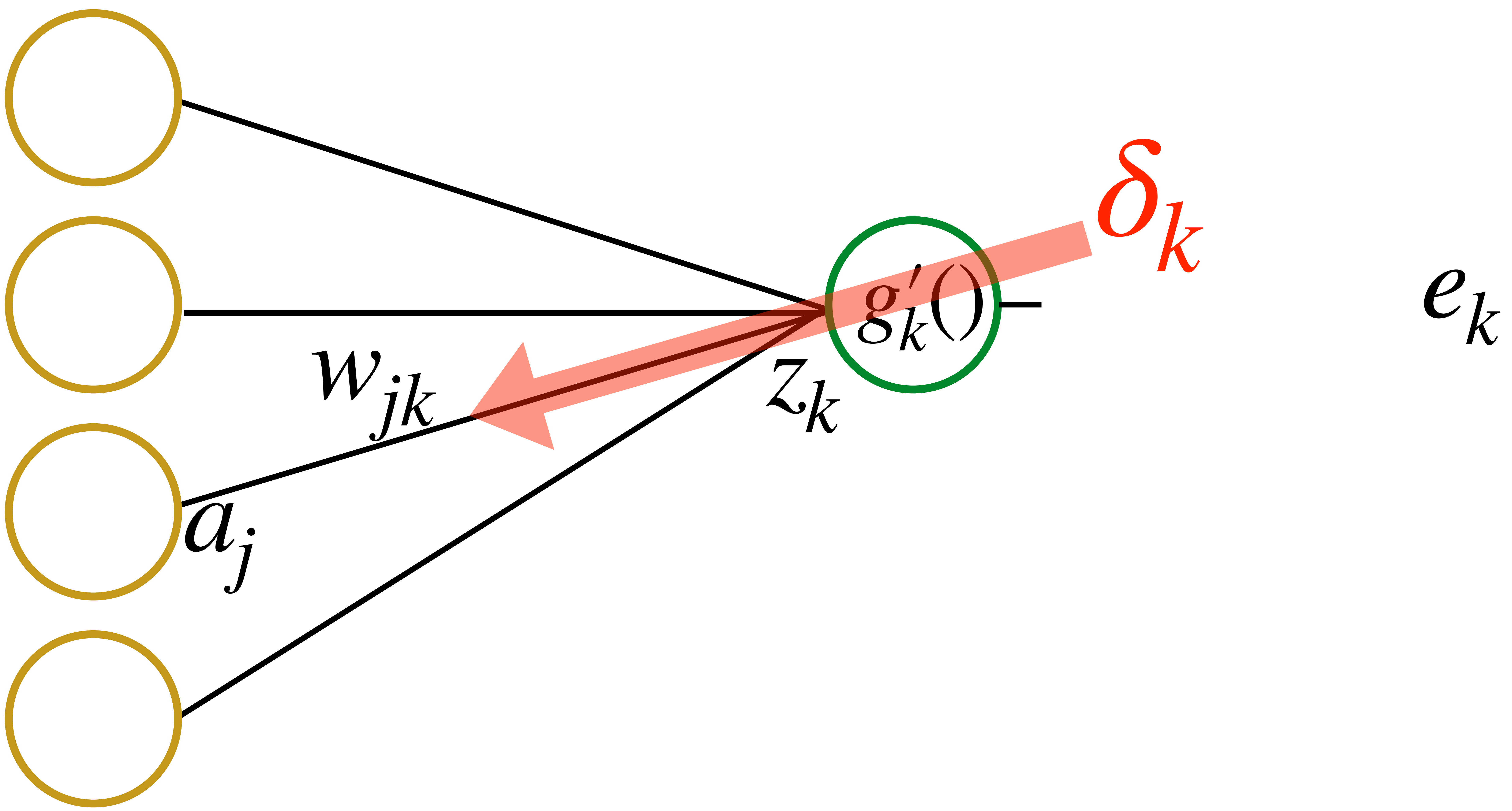
$$\frac{\partial E}{\partial w_{jk}} = e_k \;\; g_k'(z_k) \;\; a_j$$



$$\frac{\partial E}{\partial w_{jk}} = e_k \;\; g_k'(z_k) \;\; a_j \;\; = \delta_k \, a_j$$



$$\frac{\partial E}{\partial w_{jk}} = e_k \;\; g_k'(z_k) \;\; a_j \;\; = \delta_k \, a_j$$



Weight update

$$\frac{\partial E}{\partial w_{jk}} =$$ how much (and in what direction) E would change in response to a small change in w_{jk}

so <u>make</u> a small change in w_{jk} that causes E to get a little bit smaller

Weight update

$$\frac{\partial E}{\partial w_{jk}} =$$ how much (and in what direction) E would change in response to a small change in w_{jk}

so <u>make</u> a small change in w_{jk} that causes E to get a little bit smaller

$$w_{jk} \leftarrow w_{jk} - \eta \frac{\partial E}{\partial w_{jk}}$$