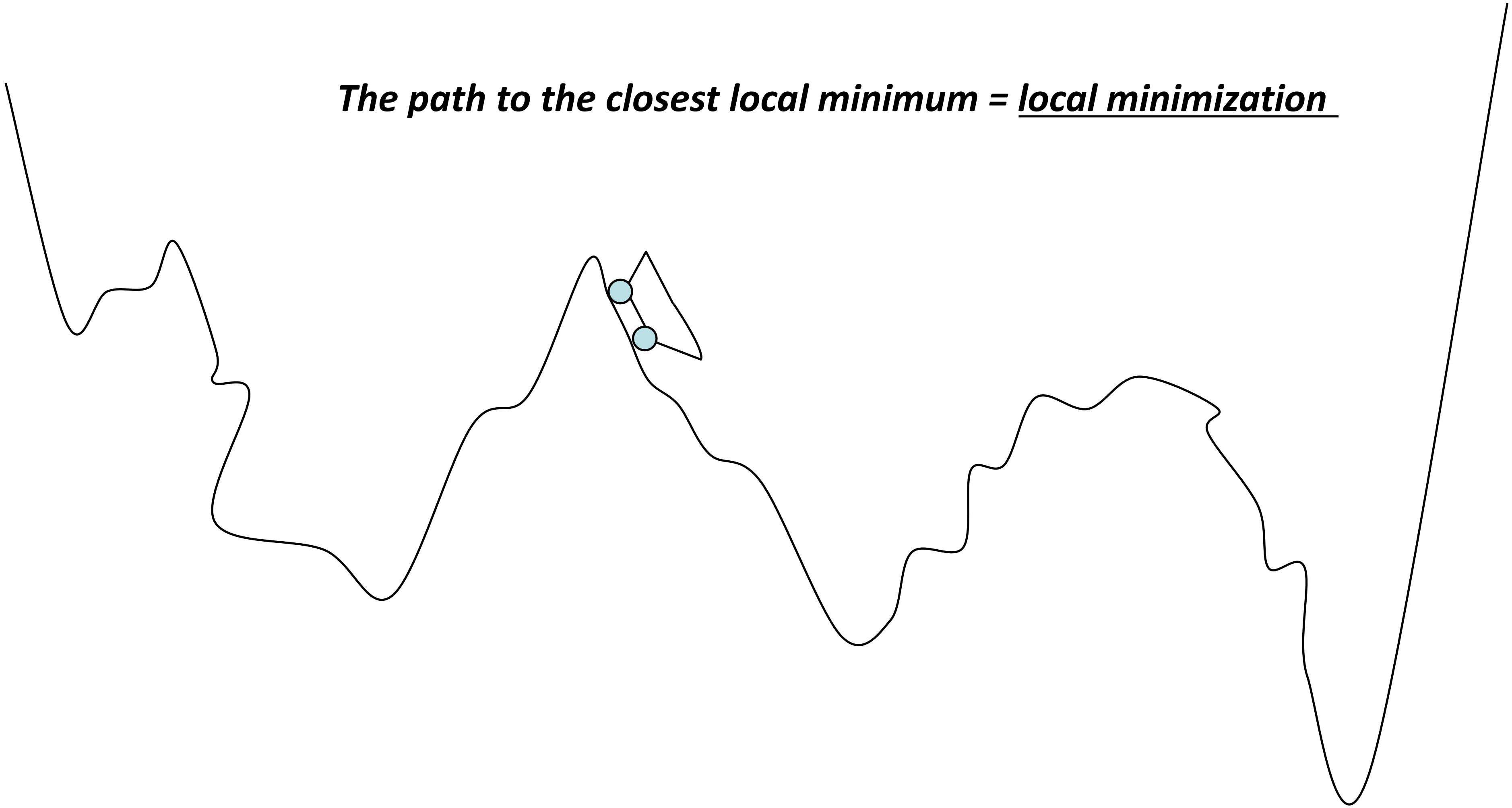
Optimization methods

Morten Nielsen
Department of Health Technology,
DTU

Minimization

CENTERFO
RBIOLOGI
CALSEQU
ENCEANA
LYSIS **CBS**

The path to the closest local minimum = <u>local minimization</u>

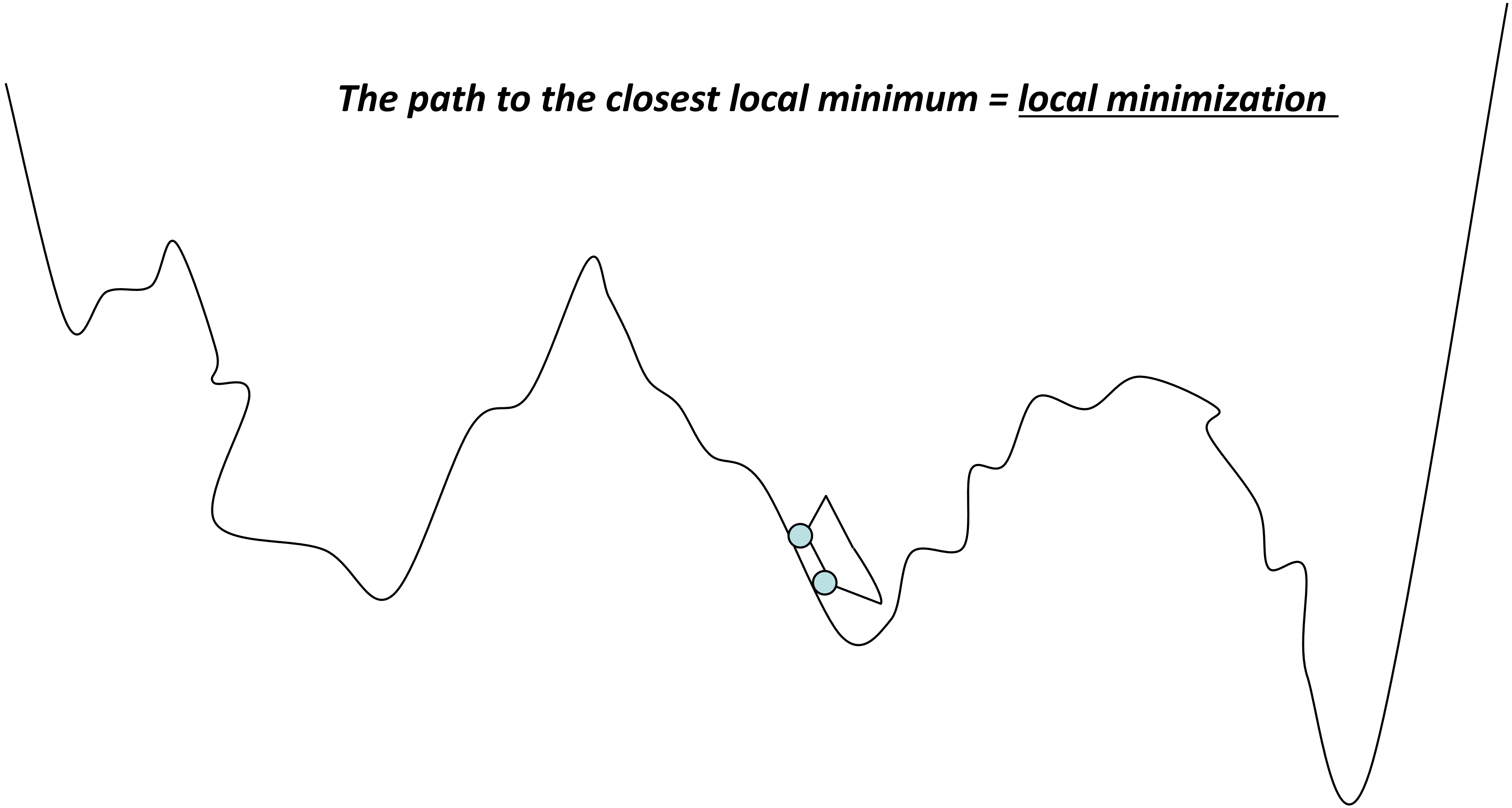


**Adapted from slides by Chen Kaeasar, Ben-Gurion University*

Minimization

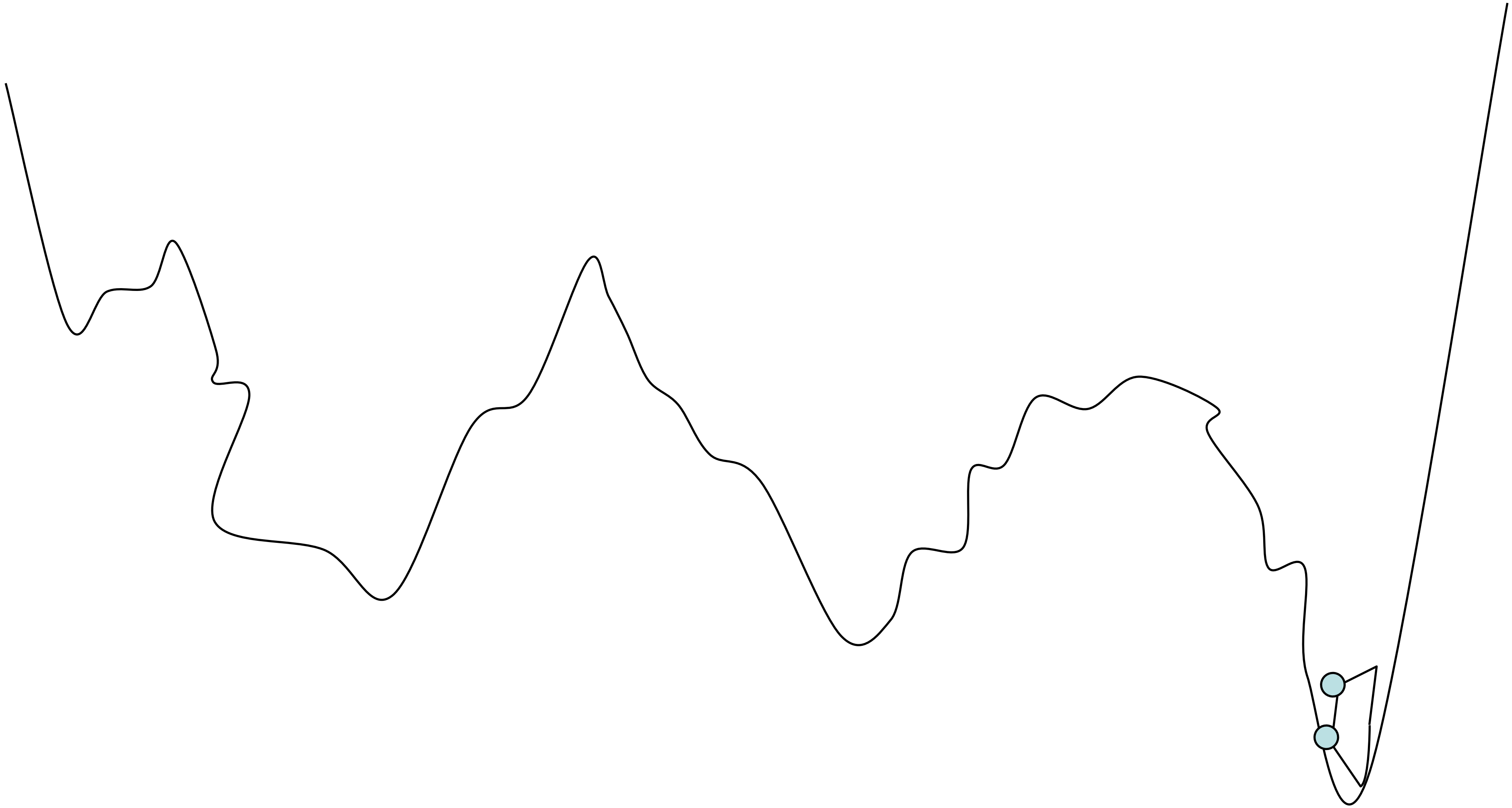
CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS **CBS**

The path to the closest local minimum = <u>local minimization</u>



**Adapted from slides by Chen Kaeasar, Ben-Gurion University*

Minimization



The path to the global minimum

Outline

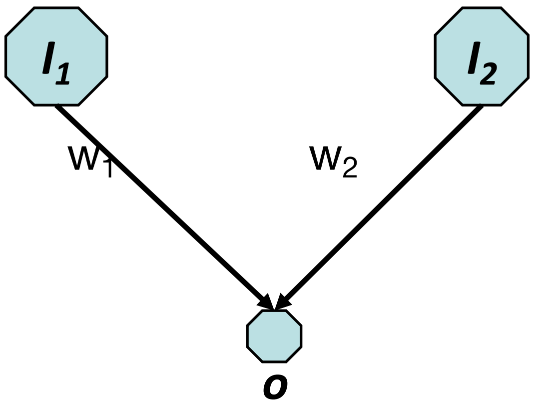
- Optimization procedures
  - Gradient descent
  - Monte Carlo

Linear methods. Error estimate

Linear function

$$o = I_1 \cdot w_1 + I_2 \cdot w_2$$

$$E = \tfrac{1}{2} \cdot (o - t)^2$$



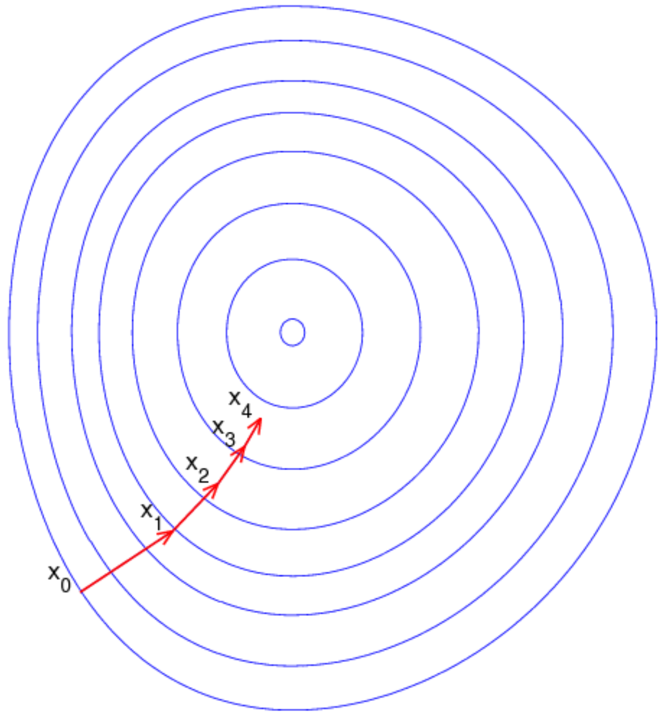
Gradient descent (from wekipedia)

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

Gradient descent is based on the observation that if the real-valued function F(x) is defined and differentiable in a neighborhood of a point **a**, then F(x) decreases fastest if one goes from **a** in the direction of the negative gradient of F at **a**.
It follows that, if

$$b = a - \varepsilon \cdot \nabla F(a)$$

for $\varepsilon > 0$ a small enough number, then $F(b) < F(a)$



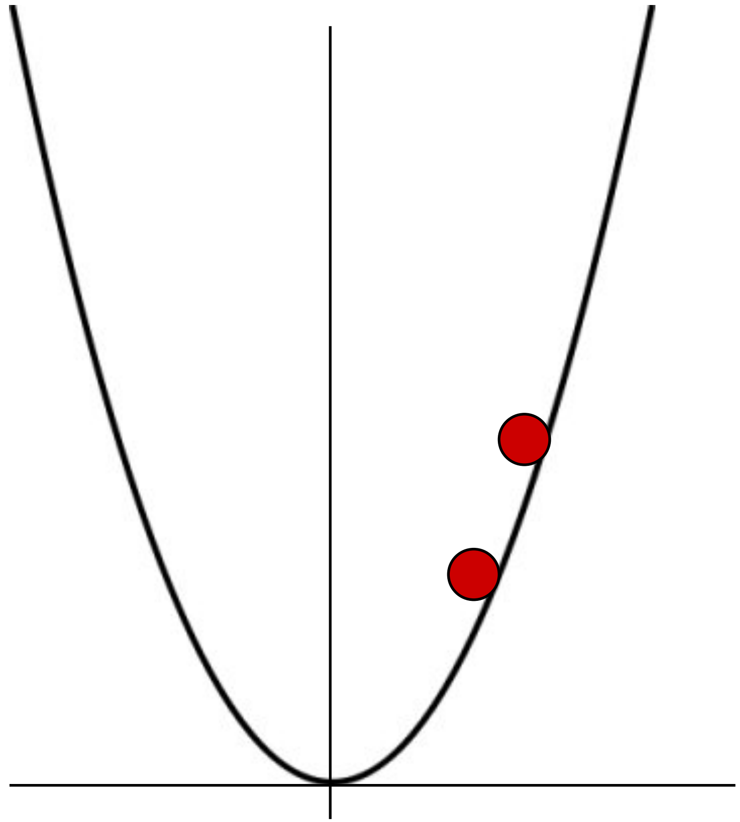
Gradient descent (example)

$$F(x) = x^2$$

$$\frac{\partial F}{\partial x} = 2 \cdot x$$

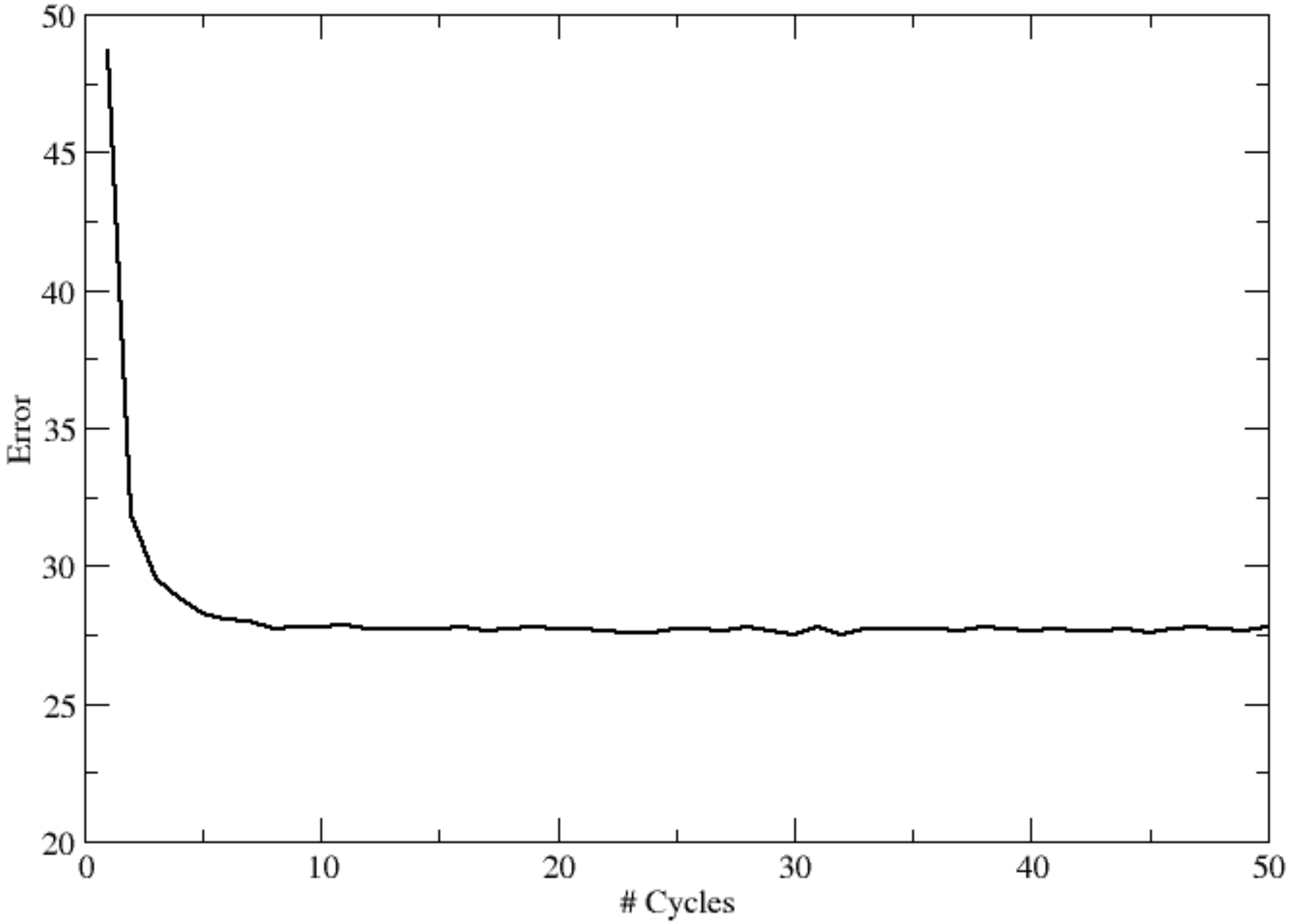
$$a = 2$$

$$F(a) = 4$$



$$b = a - \varepsilon \cdot \nabla F(a) = 2 - 0.1 \cdot 2 \cdot 2 = 1.6$$

Gradient descent



Gradient descent

Weights are changed in the opposite direction of the gradient of the error

$$w_i' = w_i + \Delta w_i$$

$$\Delta w_i = -\varepsilon \cdot \frac{\partial E}{\partial w_i}$$

$$E = \tfrac{1}{2} \cdot (O - t)^2$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial w_i}$$

$$\frac{\partial E}{\partial O} = (O - t)$$

$$\frac{\partial O}{\partial w_i} = ?$$

Gradient descent (Linear function)

CENTERFORBIOLOGICALSEQUENCEANALYSIS CBS

Weights are changed in the opposite direction of the gradient of the error

$$E = \tfrac{1}{2} \cdot (O - t)^2$$

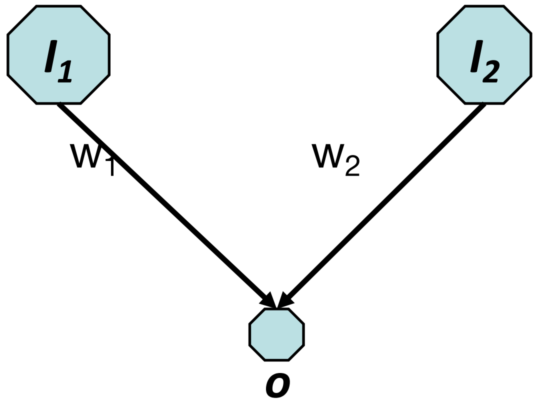
$$O = \sum_i I_i \cdot w_i$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial w_i}$$

$$\frac{\partial E}{\partial w_i} = (O - t) \cdot \frac{\partial O}{\partial w_i} = ?$$

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$



Gradient descent

Weights are changed in the opposite direction of the gradient of the error

$$E = \tfrac{1}{2} \cdot (O - t)^2$$

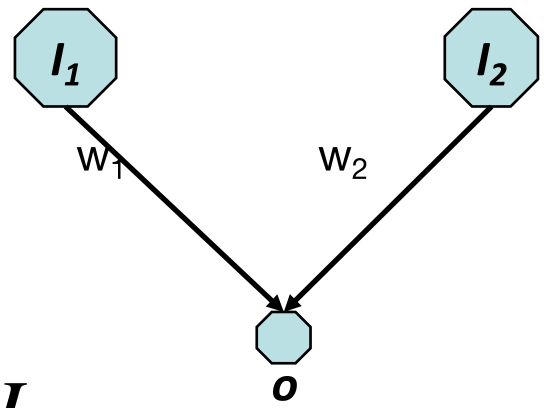
$$O = \sum_i I_i \cdot w_i$$

$$\frac{\partial E}{\partial w_i} = \frac{\partial E}{\partial O} \cdot \frac{\partial O}{\partial w_i}$$

$$\frac{\partial E}{\partial w_i} = (O - t) \cdot \frac{\partial O}{\partial w_i} = (O - t) \cdot I_i$$

Linear function

$$o = I_1 \cdot w_1 + I_2 \cdot w_2$$



Gradient descent. Example

Weights are changed in the opposite direction of the gradient of the error

$$w_i' = w_i + \Delta w_i$$

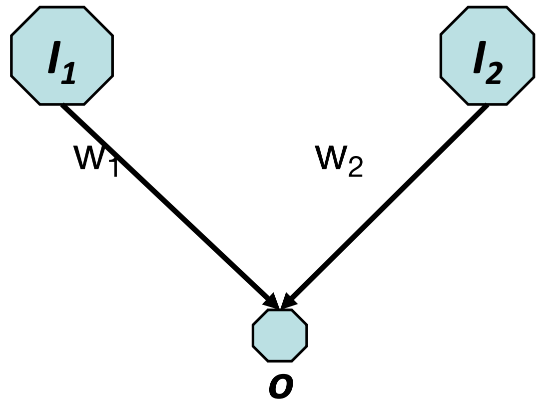
$$E = \frac{1}{2} \cdot (O - t)^2$$

$$O = \sum_i w_i \cdot I_i$$

$$\Delta w_i = -\varepsilon \cdot \frac{\partial E}{\partial w_i} = -\varepsilon \cdot ??$$

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$

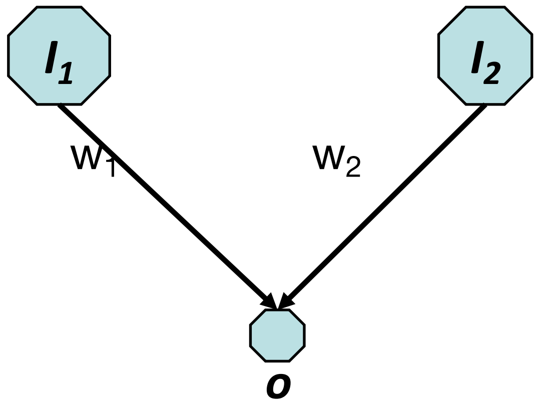


Gradient descent. Example

Weights are changed in the opposite direction of the gradient of the error

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$



$$w_i^{'} = w_i + \Delta w_i$$

$$E = \tfrac{1}{2} \cdot (O - t)^2$$

$$O = \sum_i w_i \cdot I_i$$

$$\Delta w_i = -\varepsilon \cdot \frac{\partial E}{\partial w_i} = -\varepsilon \cdot (O - t) \cdot I_i$$

Gradient descent. Doing it your self

Weights are changed in the opposite direction of the gradient of the error

$$w_i' = w_i + \Delta w_i$$

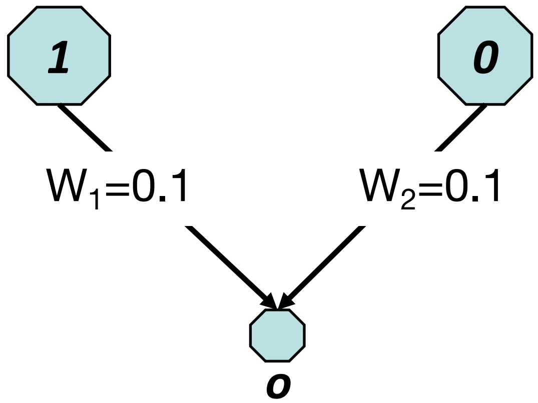
$$E = \tfrac{1}{2} \cdot (O - t)^2$$

$$O = \sum_i w_i \cdot I_i$$

$$\Delta w_i = -\varepsilon \cdot \frac{\partial E}{\partial w_i} = -\varepsilon \cdot (O - t) \cdot I_i$$

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$



What are the weights after 2 forward (calculate predictions) and backward (update weights) iterations with the given input, and has the error decrease (use ε=0.1, and t=1)?

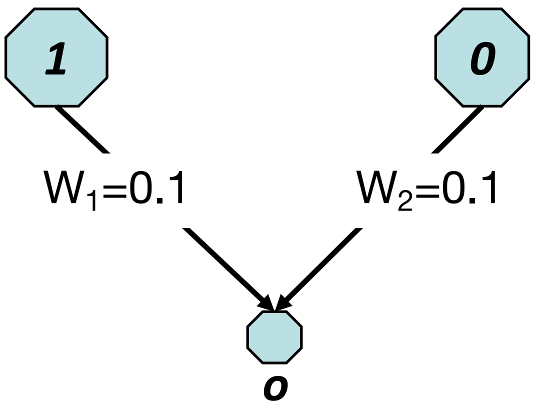
Fill out the table

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

What are the weights after 2 forward/backward iterations with the given input, and has the error decrease (use ε=0.1, t=1)?

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$



itr	W1	W2	O
0	0.1	0.1	
1			
2			

Fill out the table

Data redundancy reduction algorithms (Hobohm1 and Hobohm2). [PDF].

10.00 - 10.45 "Recorded"
Optimization procedures - Gradient decent, Monte Carlo
Optimization procedures [PDF]
GD handout

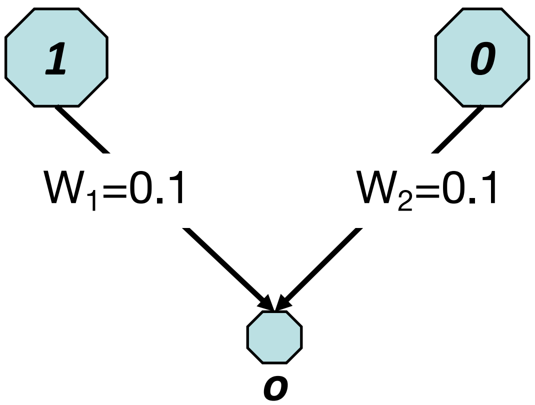
10.45 - 11.00
Break

Fill out the table

What are the weights after 2 forward/backward iterations with the given input, and has the error decrease (use ε=0.1, t=1)?

Linear function

$$O = I_1 \cdot w_1 + I_2 \cdot w_2$$



itr	W1	W2	O
0	0.1	0.1	0.1
1	0.19	0.1	0.19
2	0.27	0.1	0.27