Lecture FYS-Stk3155/4155, September 30, 2024

Basic algo : GD, and GD with momentum

GD with momentum

Define learning rate γ

Initialize β-parameters

Set # of iterations, set $\delta=[0,1]$

For # iterations and while not reached stop criterion

$$v_{(i)} = \delta(P_{(i)} - P_{(i-1)}) - \gamma g(P_{(i)})$$

(no momentum $-\gamma g(P_{(i)})$)

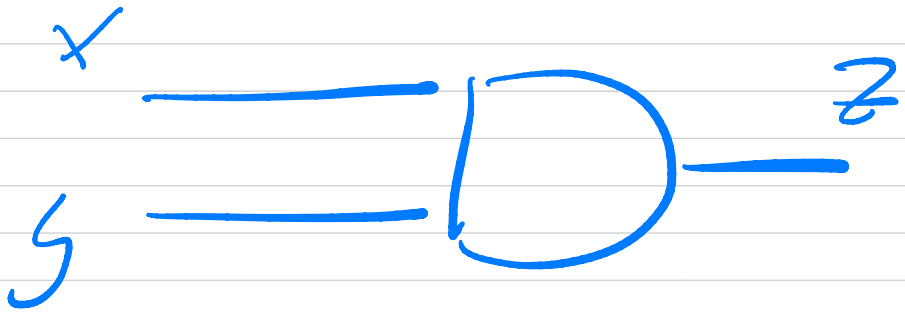
$$\beta_{(i+1)} = \beta_{(i)} + \sigma_i^{-1}$$

stop if $|\beta_{(i+1)} - \beta_{(i)}| \leq \varepsilon$

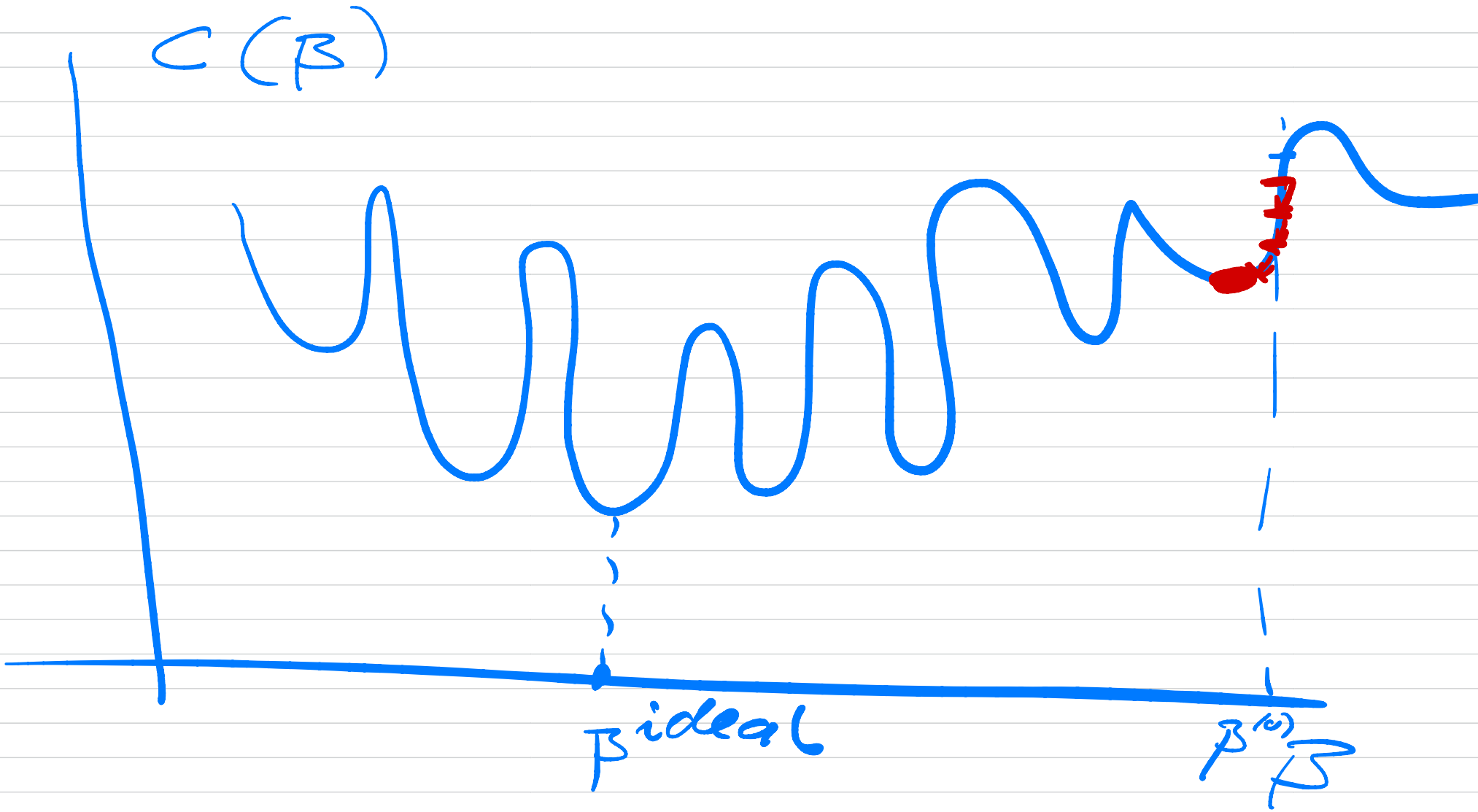
$\varepsilon = 10^{-5}$

end while.

AND - GATE



x	y	z
0	0	0
0	1	0
1	0	0
1	1	1


C(β)
β ideal
β (c)

Updating learning rate

- constant γ
- exponential decay

$$\gamma^{(k)} = \gamma^{(0)} \exp(-k\gamma_r)$$

γ_r is a parameter

- linear ↓ initial value

$$\gamma^{(k)} = (1-\alpha)\gamma^{(0)} + \alpha\gamma_r$$

γ_r is a parameter $\gamma_r \sim \frac{1}{100}\gamma^{(0)}$

select an initial $\gamma^{(0)}$

$$\gamma^{(0)} = [10^{-5}, 10^{-4}, \ldots 10^{1}]$$

- ADAgrad
- RMS prop
- ADAM

ADAGRAD

require learning (initial) $\gamma^{(0)}$
initial guess $\beta^{(0)}$
constant δ (small) for
numerical stability.
while stopping criterion
not met
(with SGD or without)
- compute gradient $g(\beta^{(i)})$
- accumulate squared
gradient
$r^{(i+1)} = r^{(i)} + g \cdot g(\beta^{(i)})$

compute update of

$$\gamma^{(i)} = \frac{\gamma^{(0)}}{\delta + \sqrt{r^{(i)}}}$$

update

$$\beta^{(i+1)} = \beta^{(i)} - \gamma^{(i)} g(\beta^{(i)})$$

end while

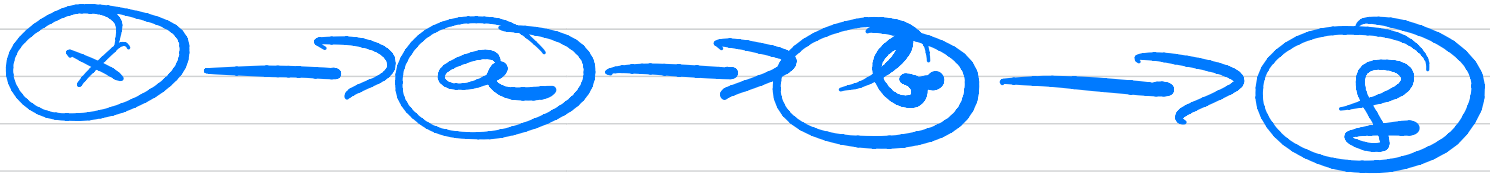
Automatic differentiation

Example

$$f(x) = \exp(x^2)$$

$$= b = \exp(a)$$

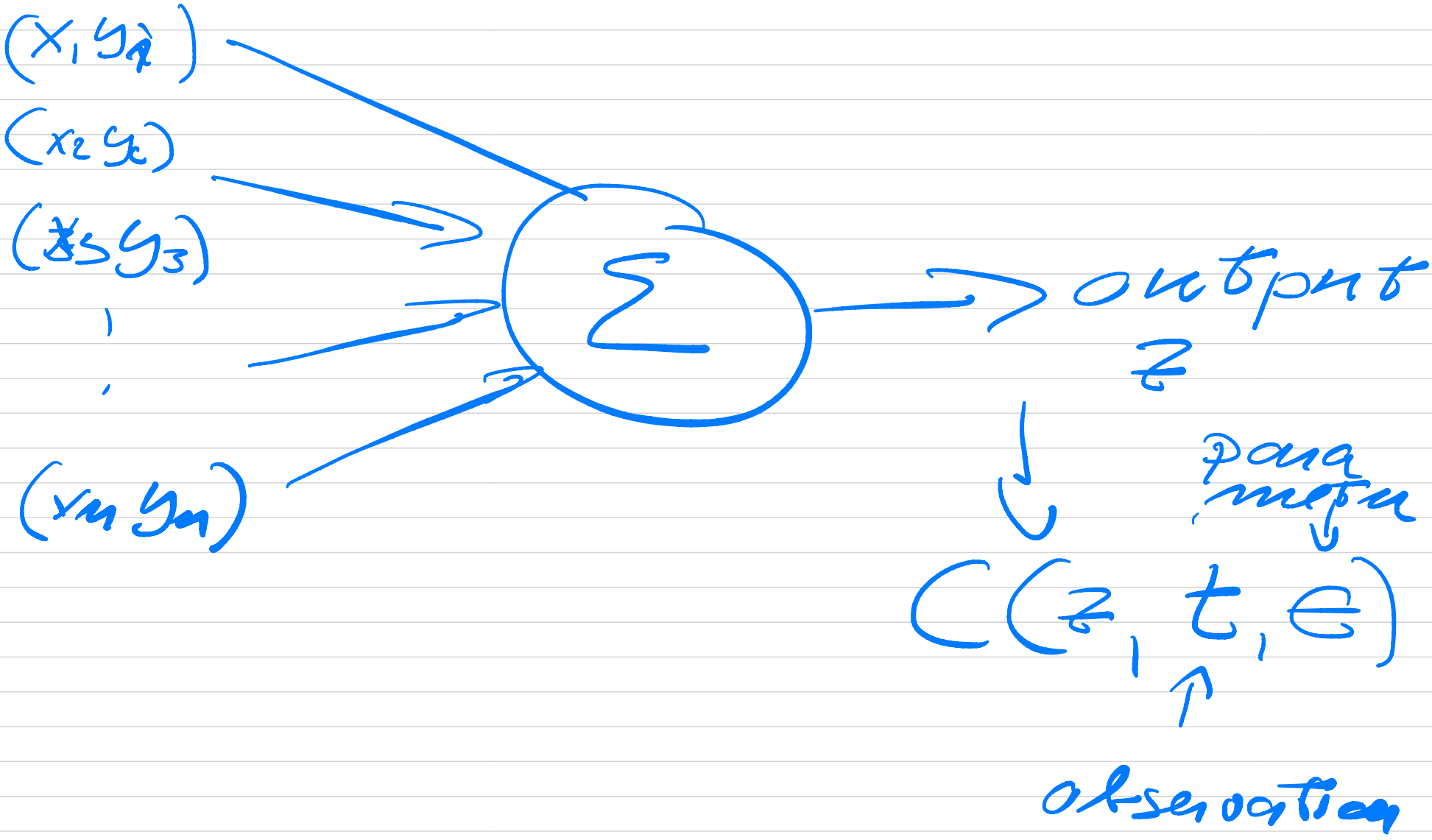
$$a = x^2$$



$$\frac{df}{dx} = \frac{df}{db}\frac{db}{da}\frac{da}{dx}$$

$$\frac{df}{db} = 1, \quad \frac{db}{da} = \exp(a), \quad \frac{da}{dx} = 2x$$

asi'de


(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
(x_n, y_n)
Σ
output
z
$C(z, t, \epsilon)$
parametre
observation

$$\frac{df}{dx} = \left[\frac{df}{db} \frac{db}{da} \right] \frac{da}{dx}$$

reverse mode

or

$$\frac{df}{db} \left[\frac{db}{da} \frac{da}{dx} \right]$$

Forward mode