Lelec2870 - Session 2

Stochastic Gradient Descent

Stochastic Gradient Descent

What is it?

Updating the weights of a model according to the gradients computed on an objective minization function

I.e. nudge the weights of a model in the right direction in order for the Loss to decrease:

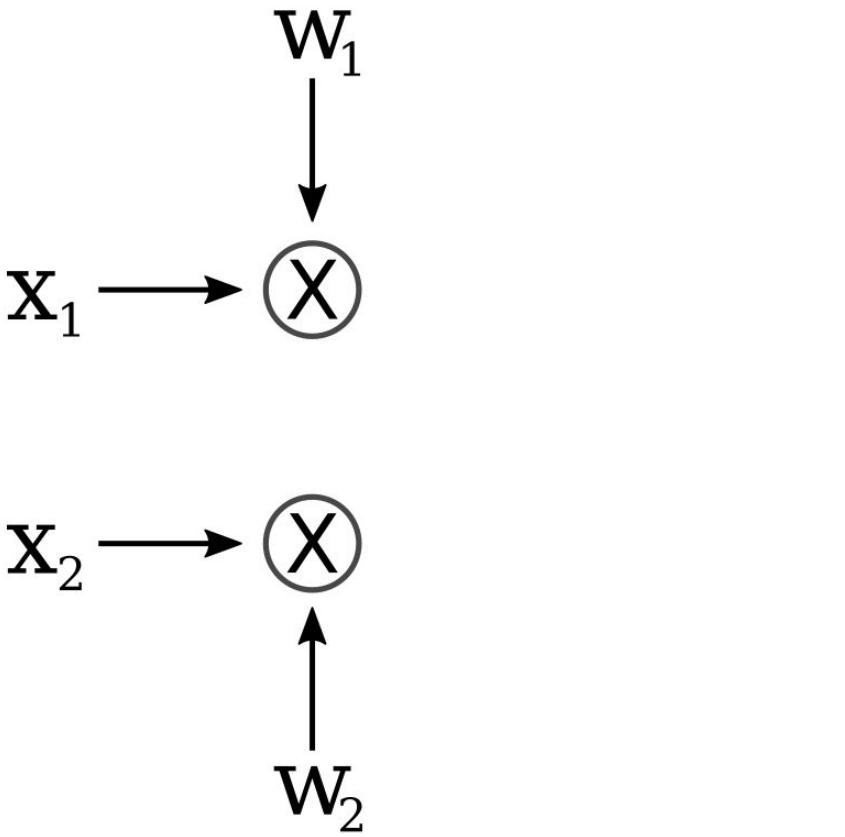
$$t_{pred} = \mathbf{wx}$$

$$\mathcal{L}(t_{pred}, t_{true}) = (t_{pred} - t_{true})^2$$

$$\mathbf{w} = \mathbf{w} - \alpha \frac{\partial \mathcal{L}}{\partial \mathbf{w}}$$

$$\iff w_i = w_i - \alpha \frac{\partial \mathcal{L}}{\partial w_i}$$

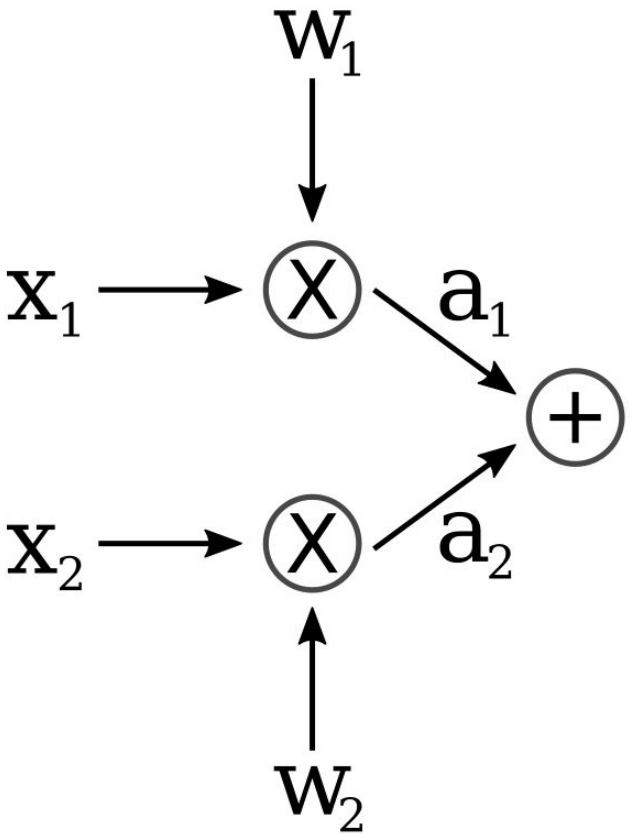
Forward Computation



x1*w1 + x2*w2 + b = t_pred

a1 = x1*w1
a2 = x2*w2

Forward Computation

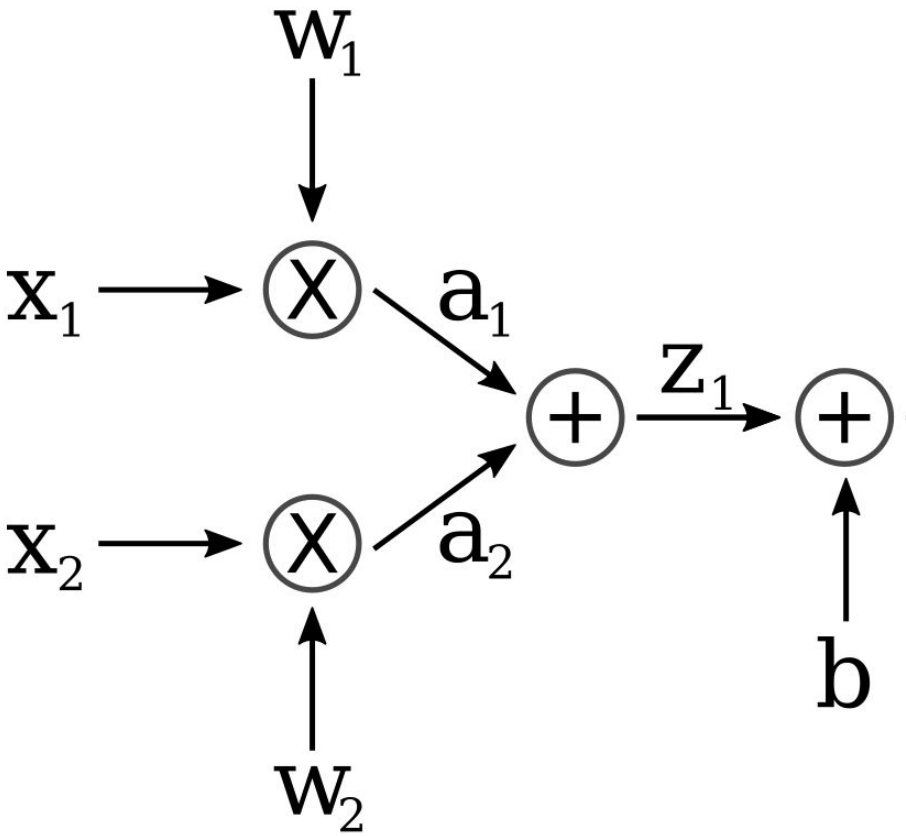


x1*w1 + x2*w2 + b = t_pred

a1 = x1*w1
a2 = x2*w2

z1 = a1 + a2

Forward Computation



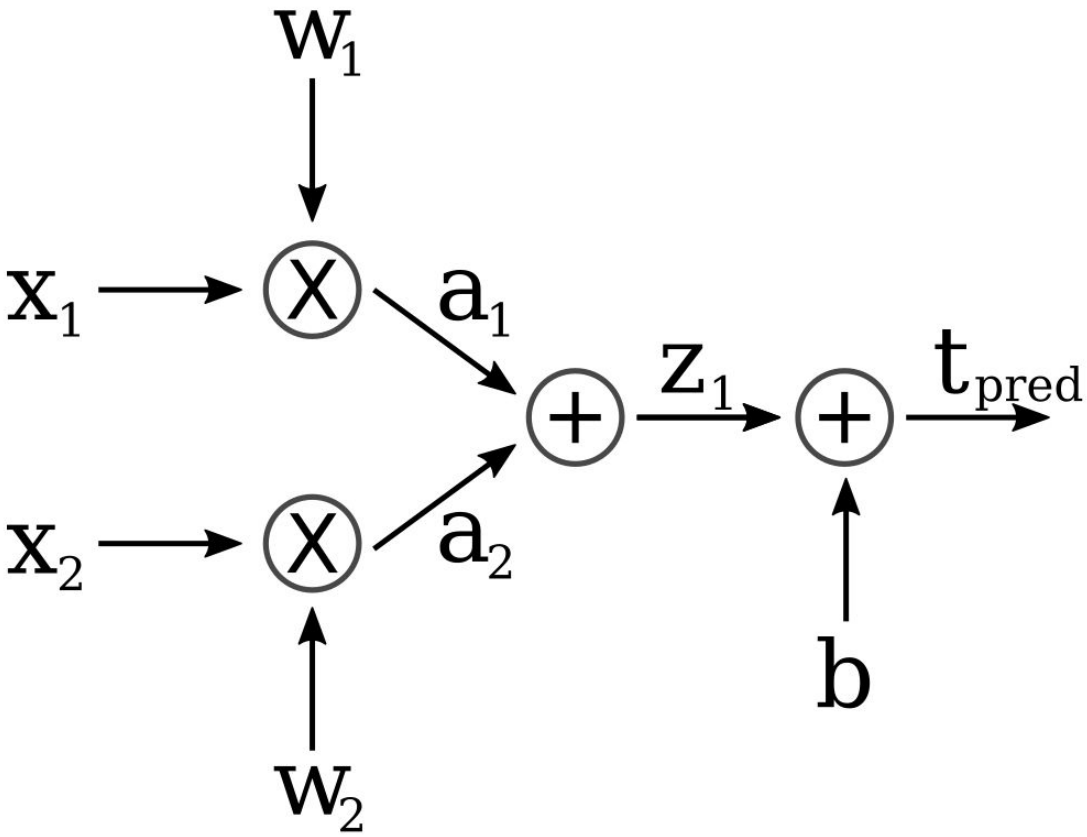
x1*w1 + x2*w2 + b = t_pred

a1 = x1*w1
a2 = x2*w2

z1 = a1 + a2

t_pred = z1 + b

Forward Computation



x1*w1 + x2*w2 + b = t_pred

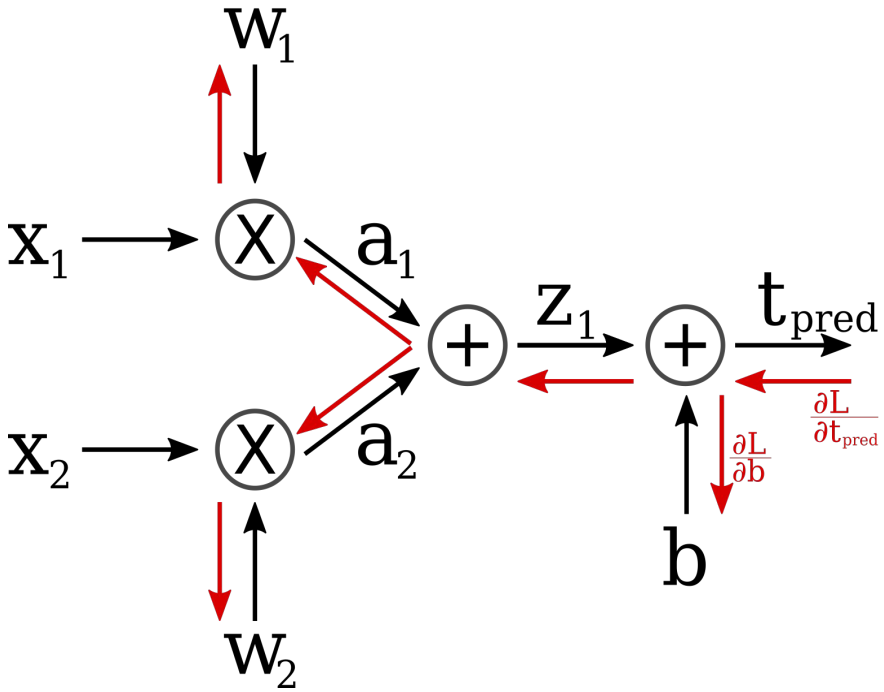
a1 = x1*w1
a2 = x2*w2

z1 = a1 + a2

t_pred = z1 + b

$L = (t\_pred - t\_true)^2$

Forward Computation



t_pred = z1 + b

L = (t_pred-t_true)²

$$L = (t_{pred} - t_{true})^2$$

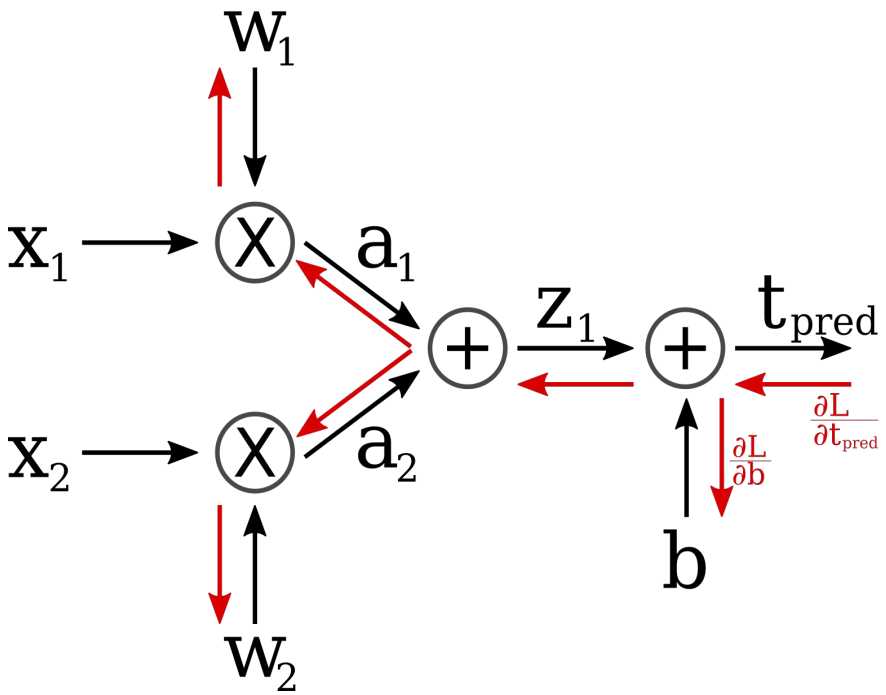
$$\frac{\partial L}{\partial t_{pred}} = 2(t_{pred} - t_{true})$$

$$\frac{\partial L}{\partial b} = \frac{\partial L}{\partial t_{pred}} \frac{\partial t_{pred}}{\partial b}$$

$$= \frac{\partial L}{\partial t_{pred}} 1$$

.
.
.

Forward Computation

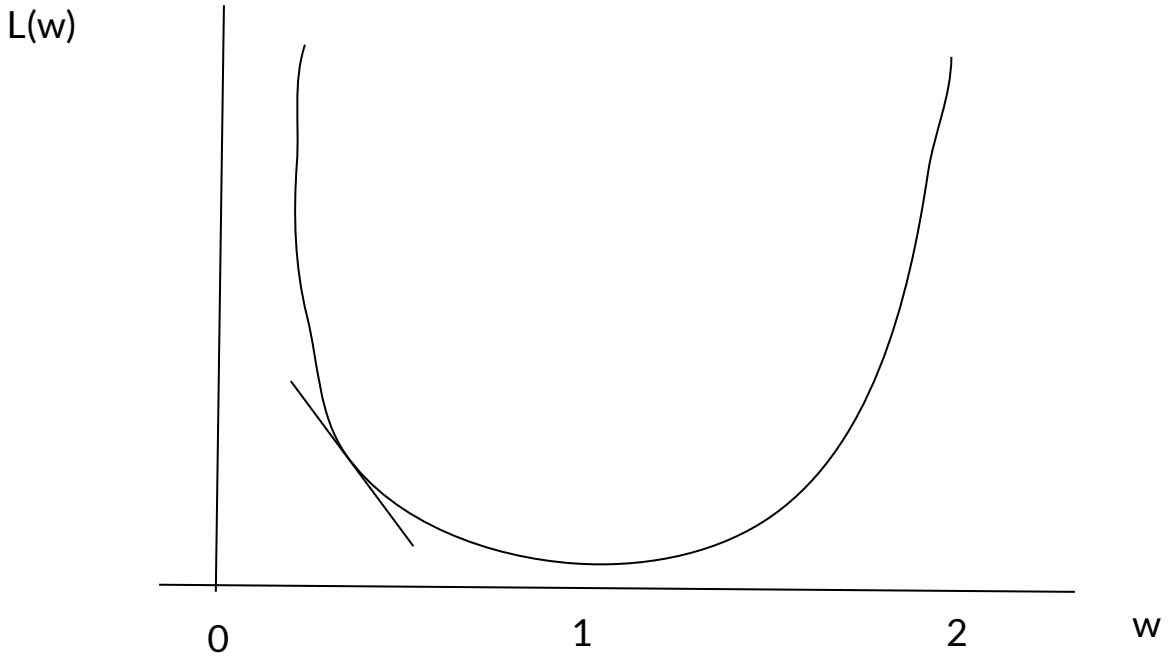


t_pred = x1*w1 + x2*w2 + b

L = (t_pred-t_true)²
t_pred = z1 + b
dL/dz1 = dL/dt*dt/dz1

dF/dt = dF/dm*dm/dt
If m(t) -> is a function of t

z1 = a1+a2
dL/da1 = dL/dz1*dz1/da1
dL/dz1*1
a1 = w1*x1
dL/dw1 = dL/da1*da1/dw1
= dL/da1*x1
w1,1 <- w1,0 - lr*dL/dw1
w1,2 <- w1,1 - lr*dL/dw1


L(w)
0
1
2
w