SGD with Momentum

Reminder from phisics:

* x^t, v^t, a^t — position, velocity and acceleration at the moment t

$$x^{t+1} = \boxed{x^t} + \boxed{\frac{1}{2} a^t t^2}$$

which looks a lot like

$$w^{t+1} = \boxed{w^t} - \boxed{\eta \frac{\partial \mathcal{L}(w^t)}{\partial w}}$$

So we can treat w^t as position

$$\begin{cases} x^{t+1} = x^t + v t \\ v^{t+1} = v^t + a^t \cdot t \end{cases}$$

Update Rule:

* $v^{t+1} = \rho v^t + \dfrac{\partial \mathcal{L}(w^t)}{\partial w}$, where

  $\rho < 1$ — momentum

* $w^{t+1} = w^t - \eta v^{t+1}$

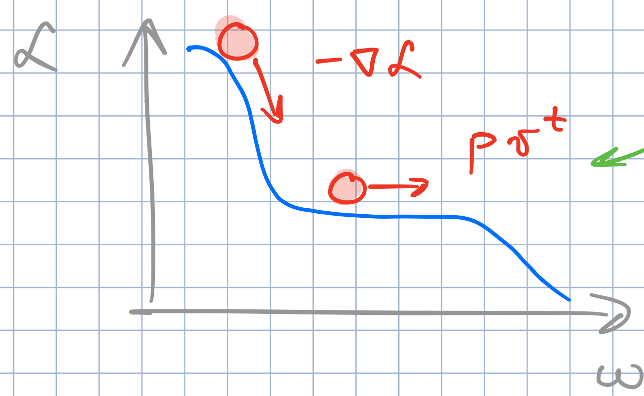
Advantages:

* avoid stucking in the settle points

  $\left(w : \dfrac{\partial \mathcal{L}(w)}{\partial w} = 0 \right)$

* avoid stucking in the local mins

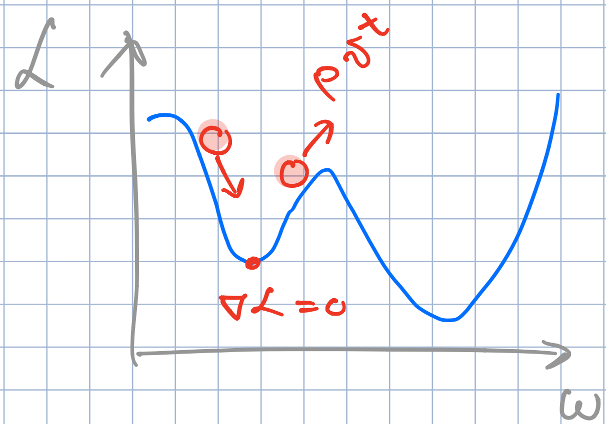
Illustration: $\measuredangle$ 1D case



Even though $\nabla\mathcal{L} = 0$ non-zero momentum ρ keep pushing ball

$$\omega^{t+1} = \omega^t - \eta(\underline{\rho}\, v^t + \nabla\mathcal{L})$$

Similar in the case of local minimum:



without $\rho > 0$

$$\omega^{t+1} = \omega^t - \eta\nabla\mathcal{L} = \omega^t$$

in the local min

i.e. we stuck!