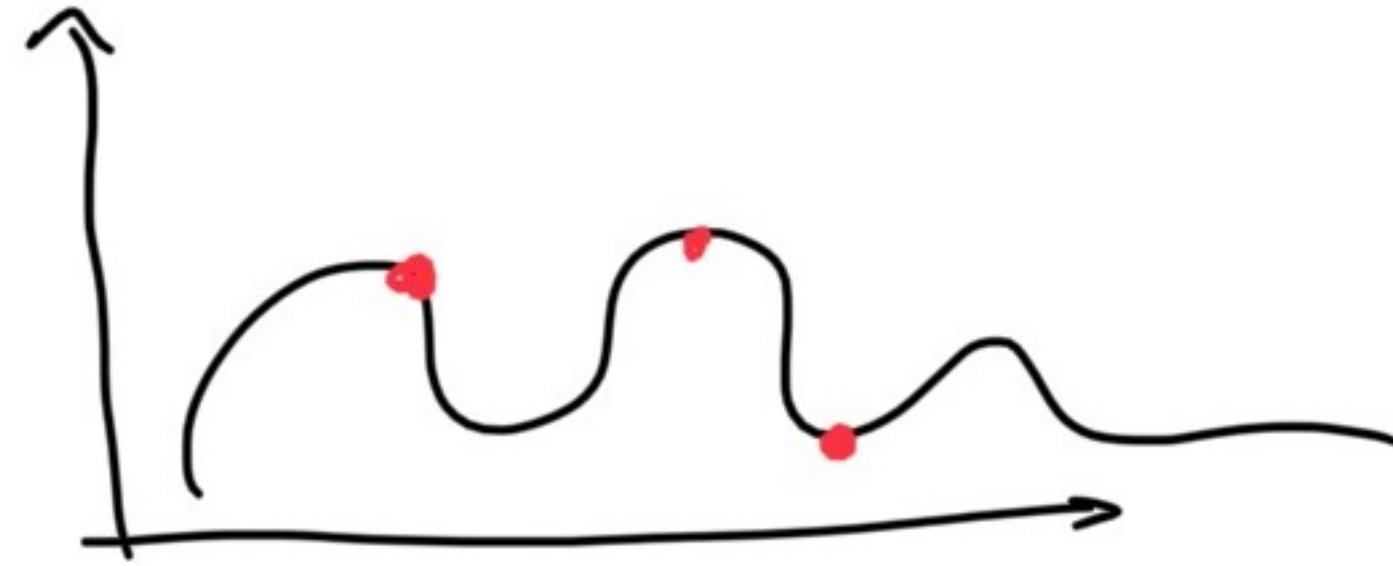


Random restart to solve local minima



fixing Vanishing Gradient problem:

1. Changing the activation function.  $\tanh$ , ReLU

Momentum to solve local minima:

$\beta$  : momentum  $\in (0, 1)$

$$\text{step}(t) = \text{step}(t) + \beta \text{step}(t-1) + \beta^2 \text{step}(t-2) \dots$$

↳ intuitively it gets us out of local minima.