$$\theta^* = \arg\min_{\Theta} \mathcal{L}(\theta, T)$$

⓪ $\theta^{(0)}$;

η - learning rate; $\quad \eta \overset{def}{=} 10^{-4}$

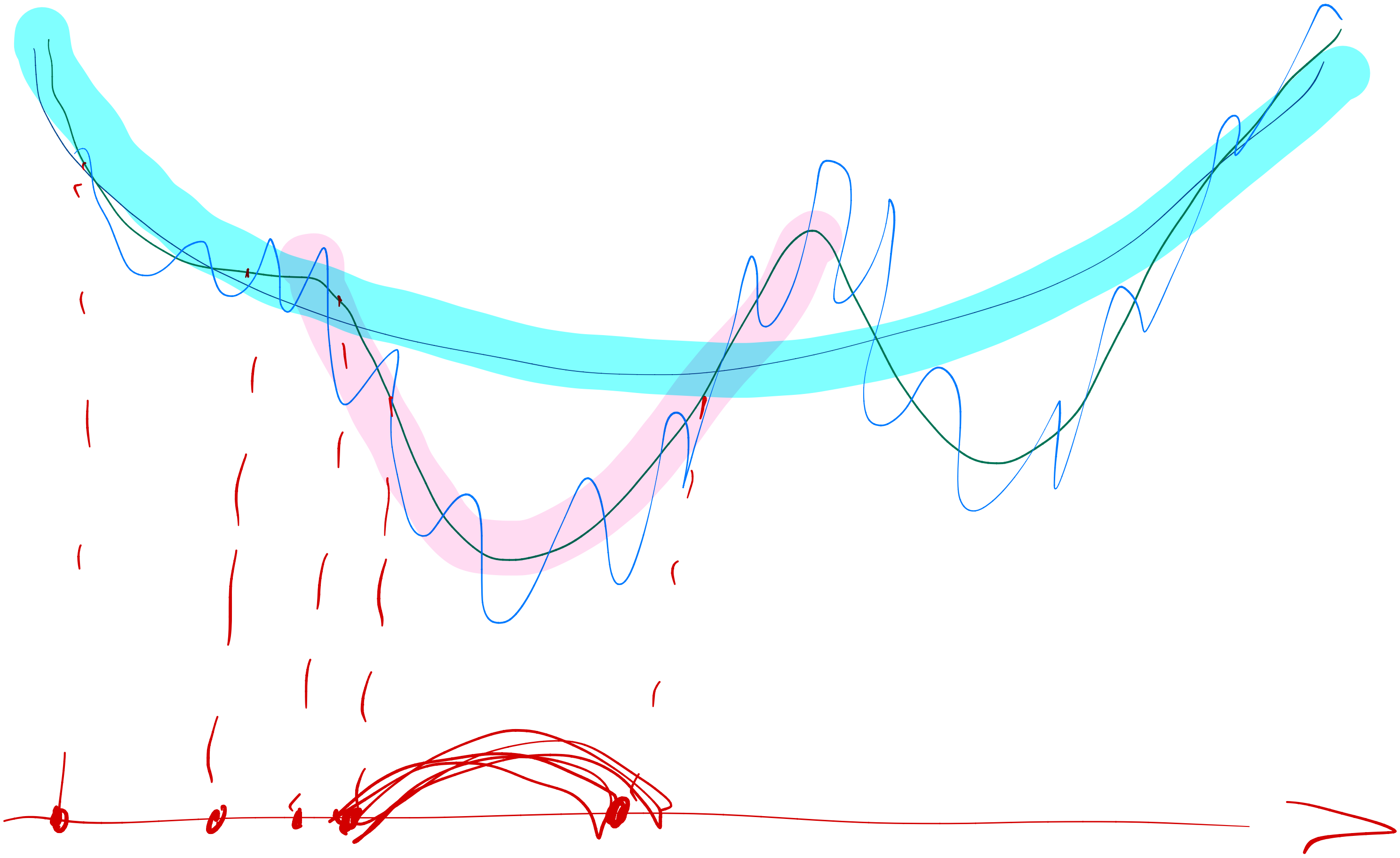
C - условие остановки.

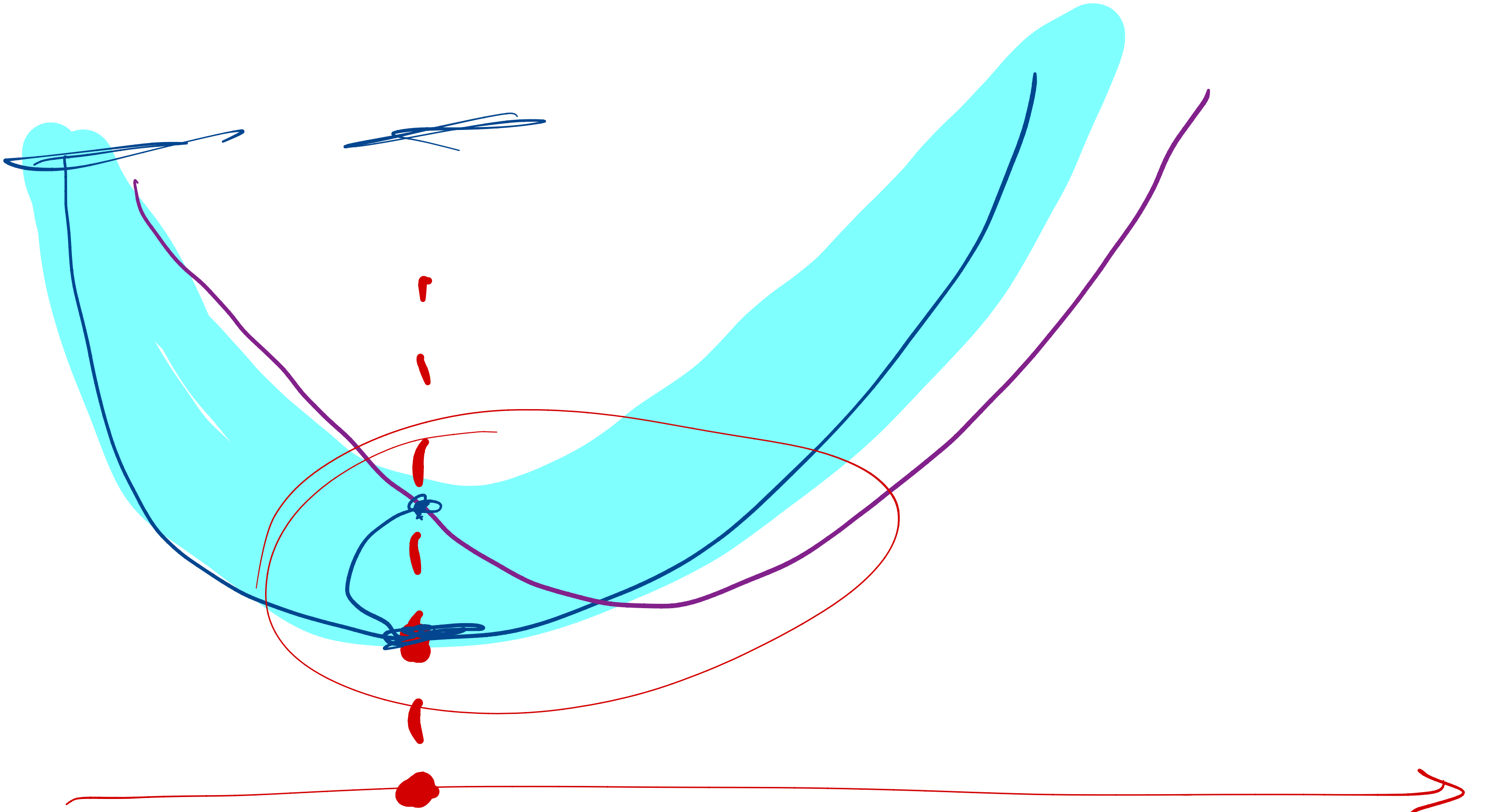
① $g_t = \nabla_\theta \mathcal{L}(\theta^{(t)}, B^{(t)})$

g^* (Adam)

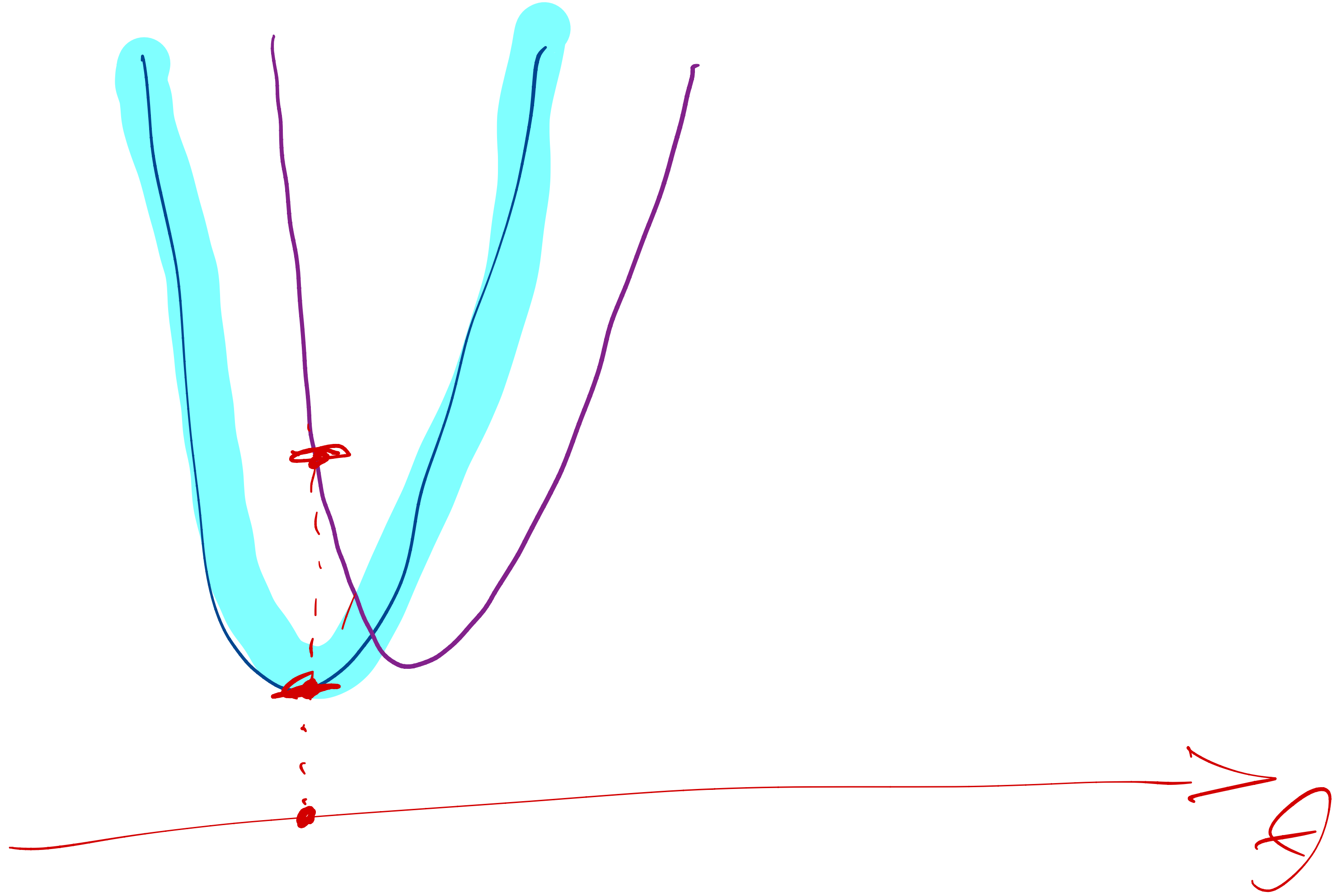
② $\theta^{(t+1)} = \theta^{(t)} - g^* \eta$

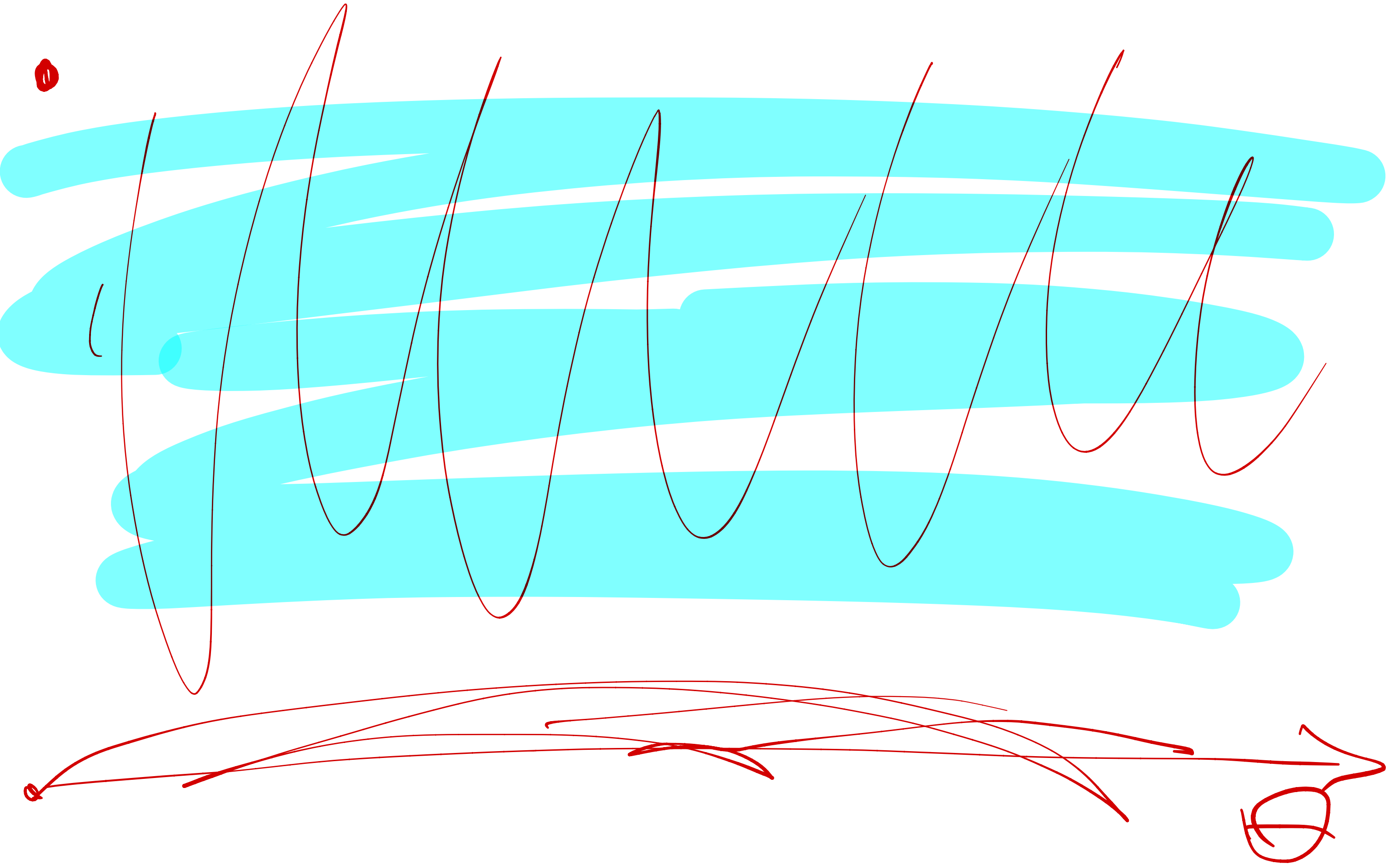
③ C? => stop

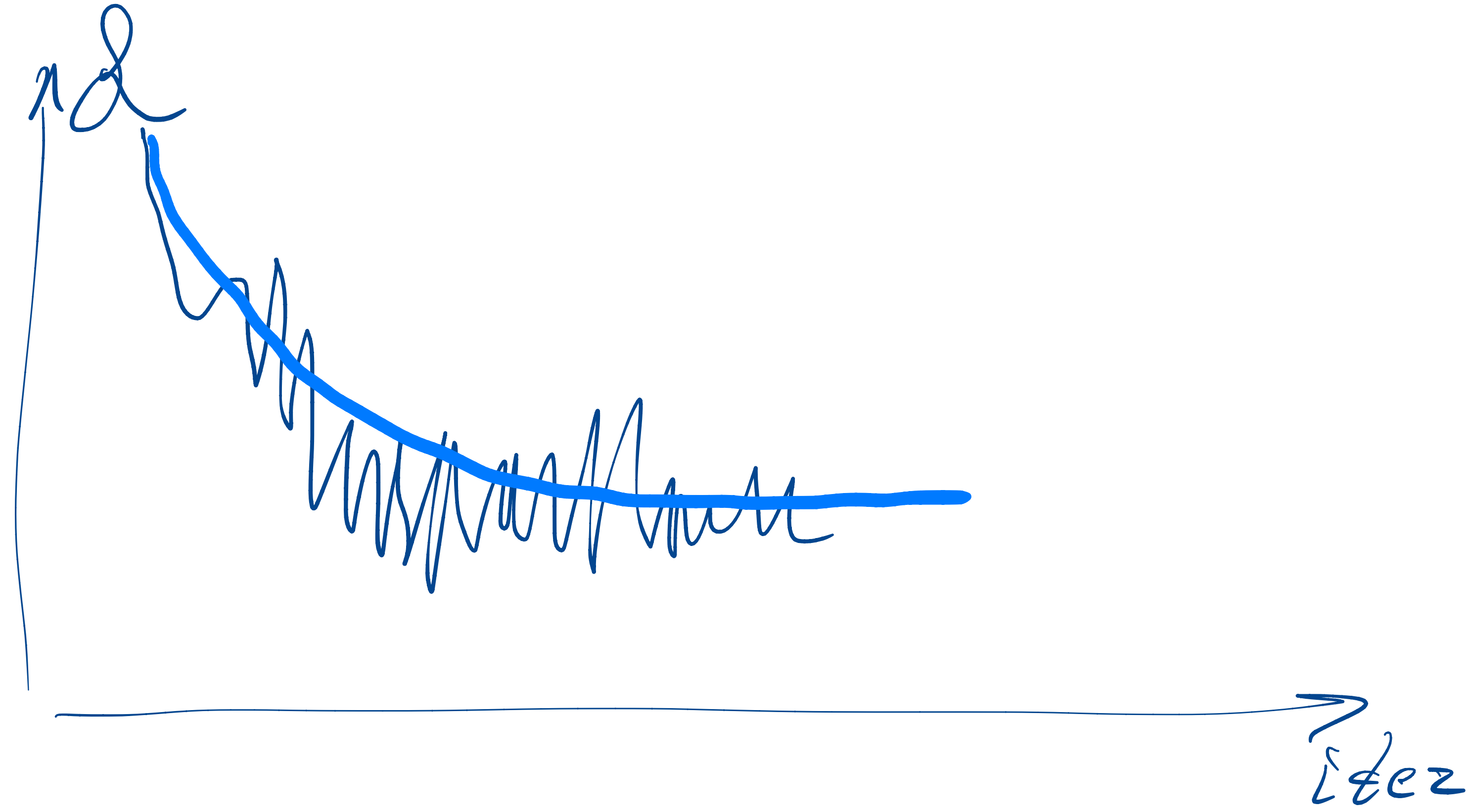




covariate shift


θ

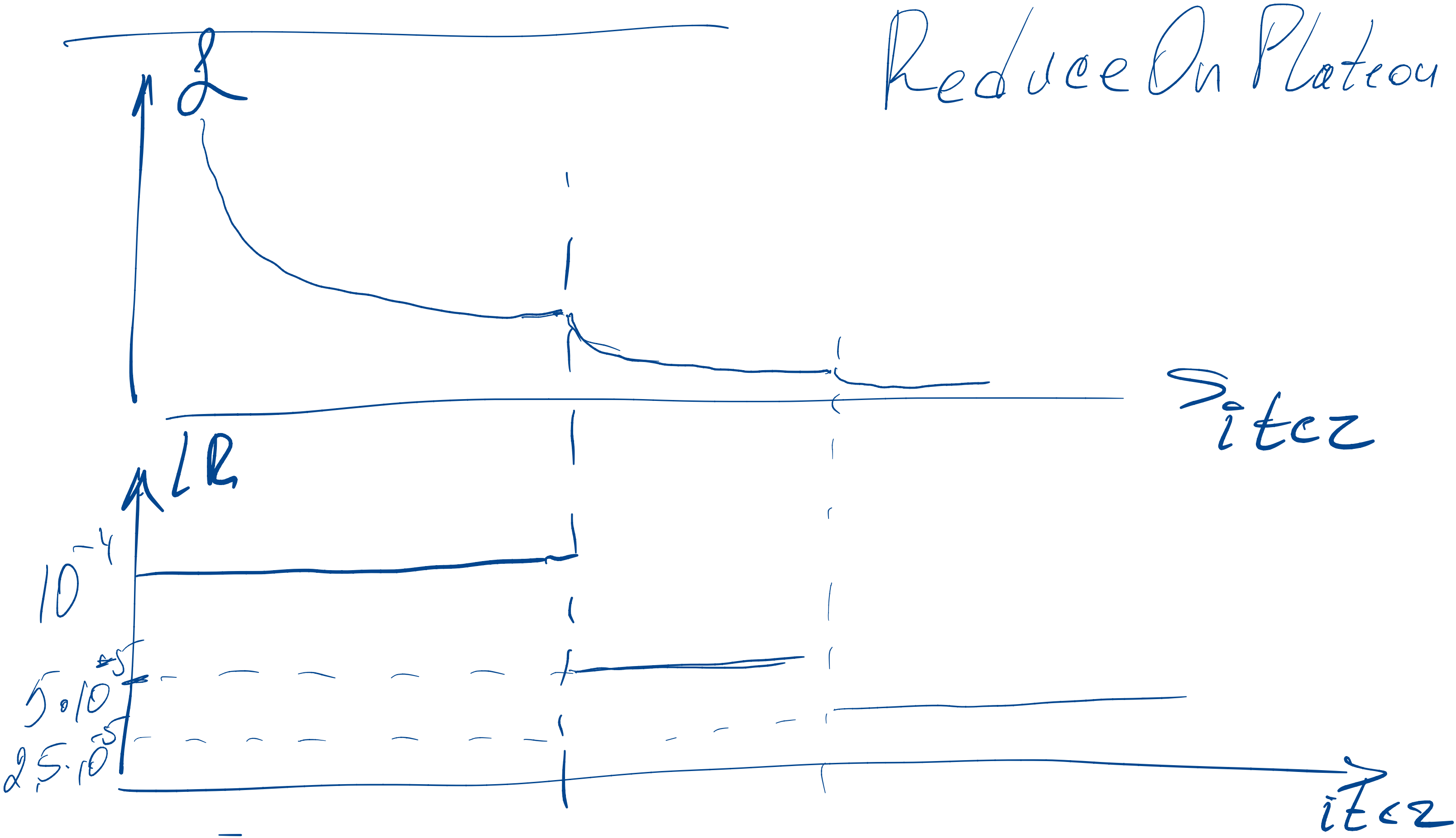

θ



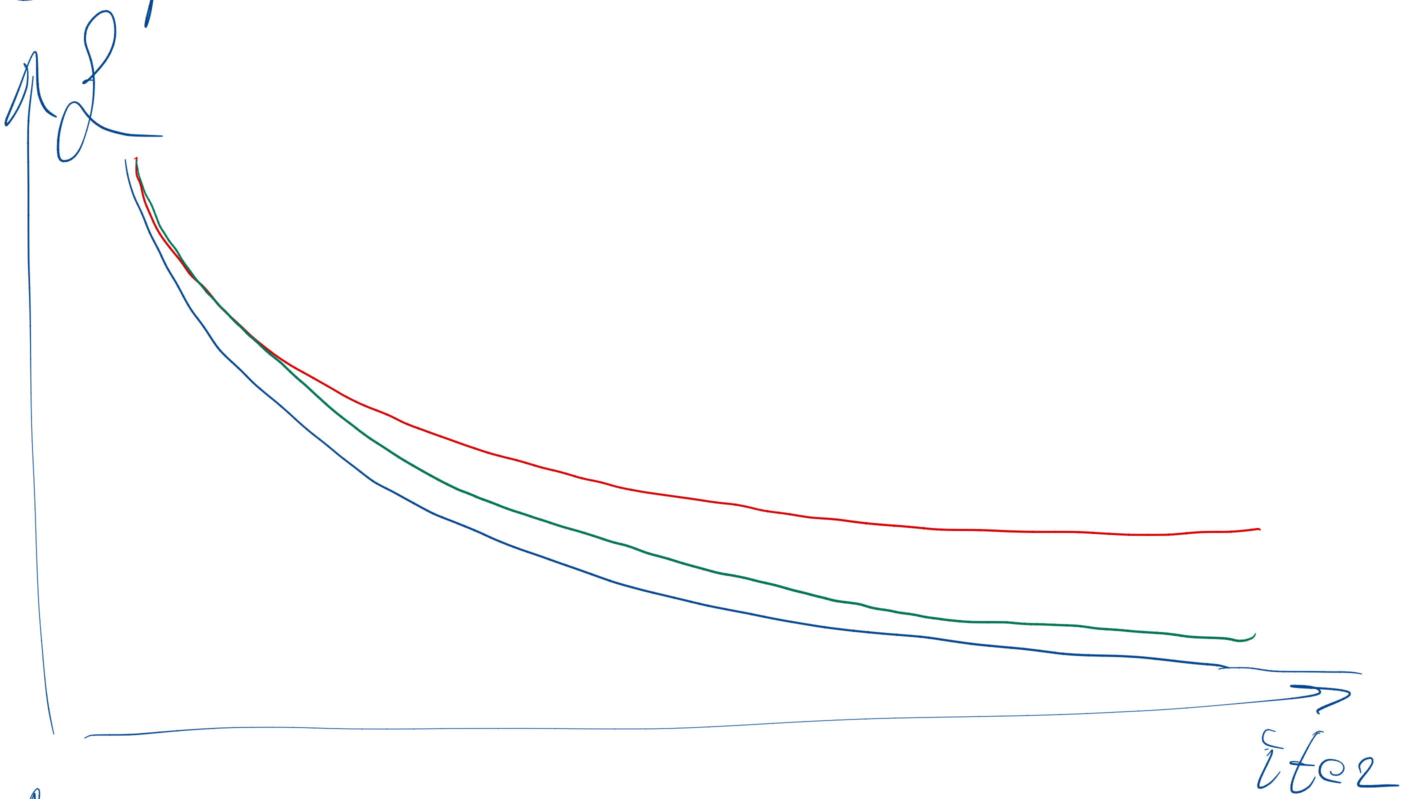
Стратегии изменения темпа обучения

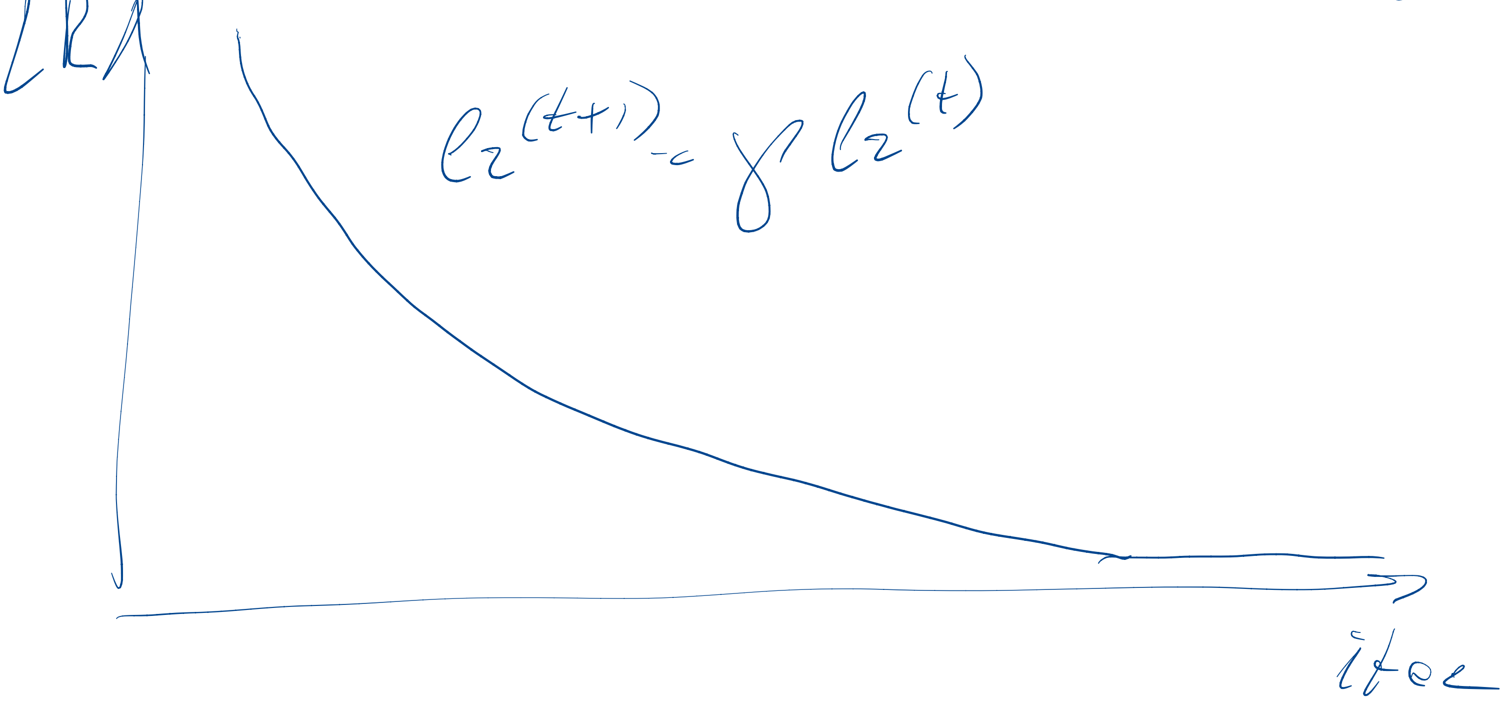
LR schedulers

ReduceOnPlateau



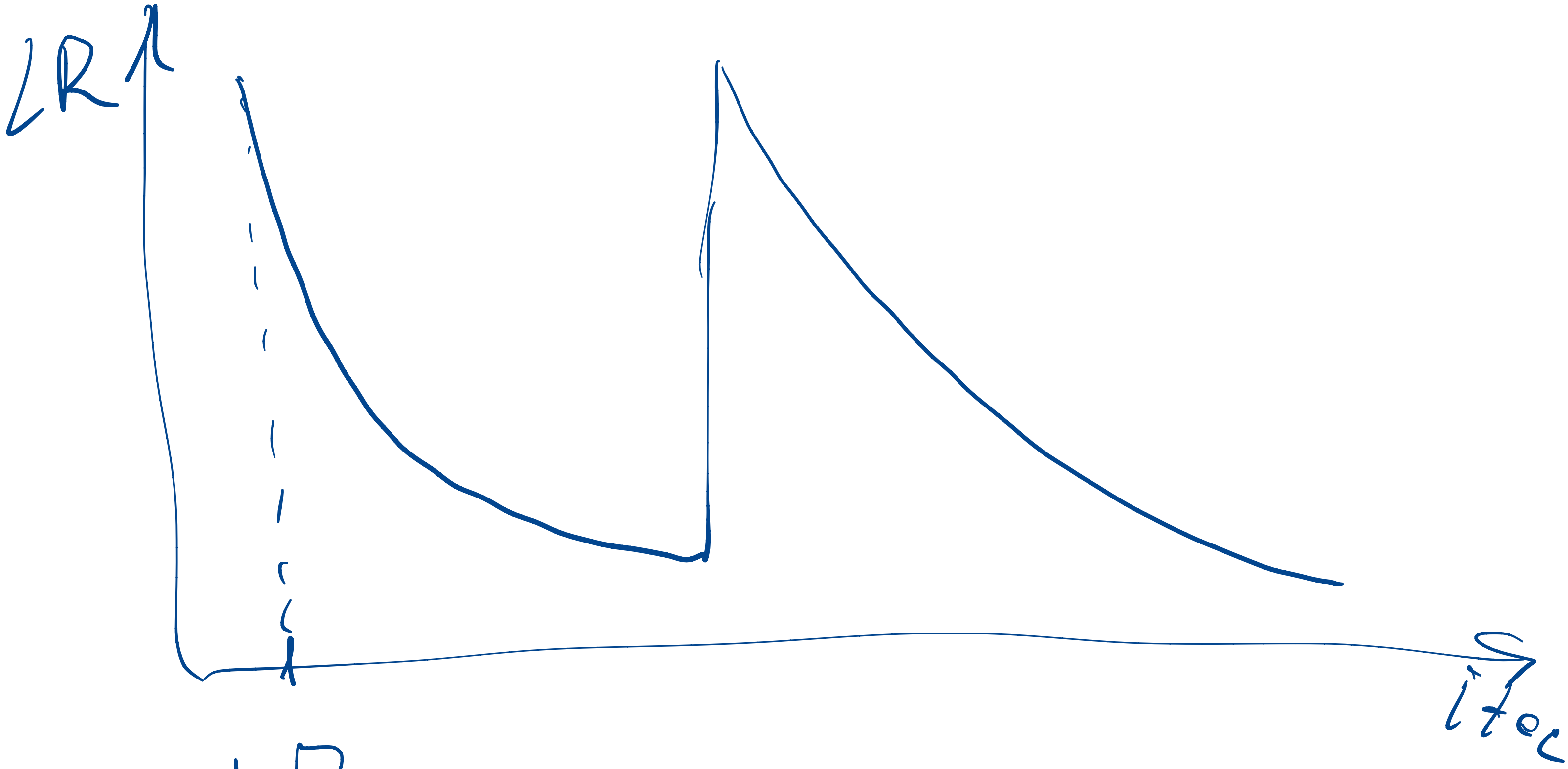
Exponential LR

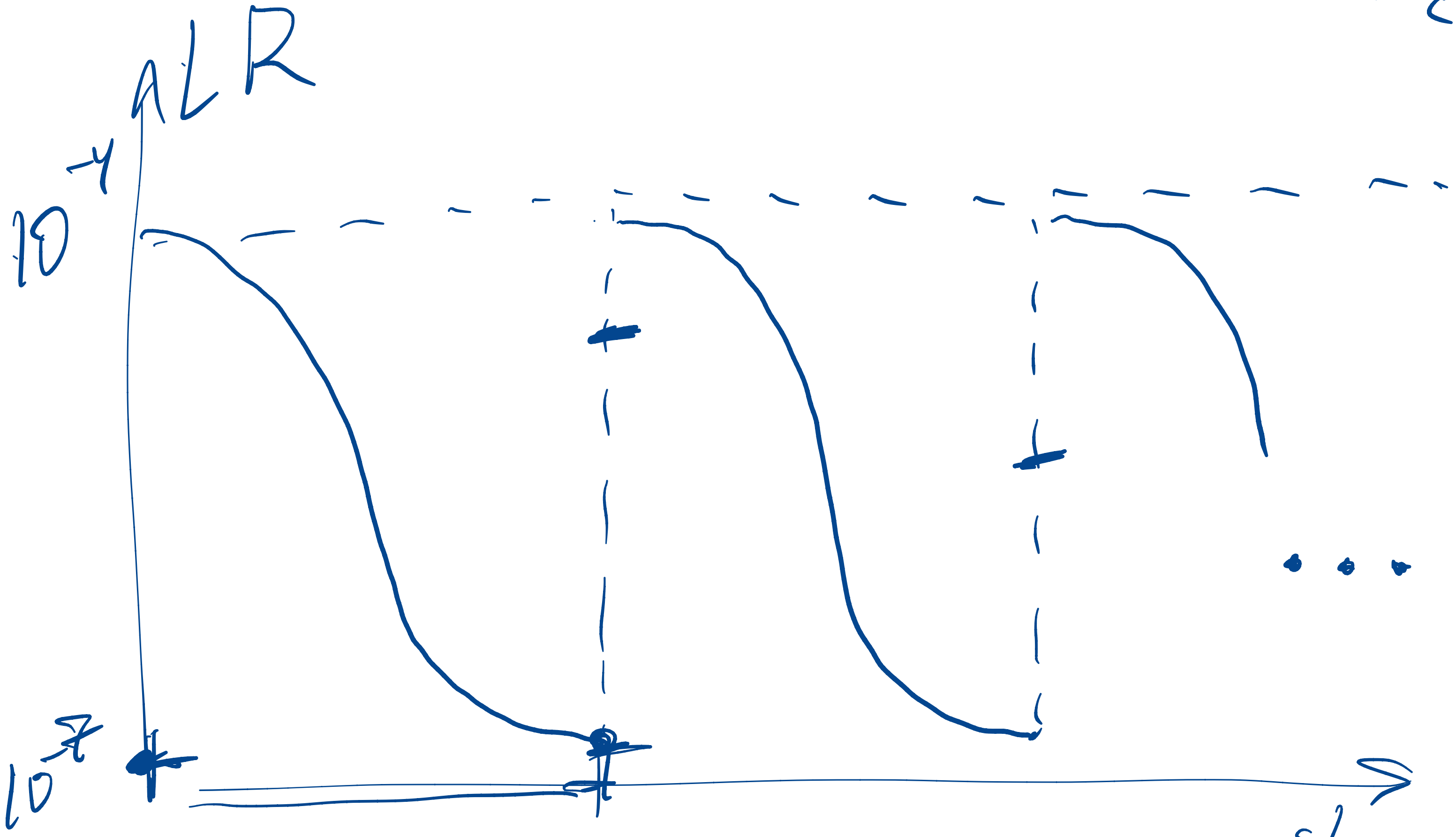




$$lr^{(t+1)} = \gamma\, lr^{(t)}$$

Simulated annealing





SGDR