Optimization

Tuesday, 3 April 2018 16:27

Xavier / MSRA Initialization.

Problem in SGD:

Zig-zag on sensitive dimension.

Saddle points will be common when dimension is large.

SGD + Momentum:

$$v_{t+1} = \rho v_t + \nabla f(x_t)$$

$$x_{t+1} = x_t - \alpha v_{t+1}$$

ρ: rub:

Nesterov Momentum:

$$v_{t+1} = \rho v_t - \alpha \nabla f(x_t + \rho v_t)$$

$$x_{t+1} = x_t + v_{t+1}$$

change of variable:

$$v_{t+1} = \rho v_t - \alpha \nabla f(\tilde{x}_t)$$

$$\tilde{x}_{t+1} = \tilde{x}_t - \rho v_t + (1+\rho) v_{t+1}$$

$$= \tilde{x}_t + v_{t+1} + \rho(v_{t+1} - v_t)$$

this algorithm will ignore sharp minima

AdaGrad: not good for non-convex

RMS Prop: