Lecture 24: Deep neural networks

Professor Ilias Bilionis

The stochastic gradient decent and variants



The Robbins-Monro

algorithm - Stochastic optimization algorithm

wjin [Fz[(19;Z)](*)

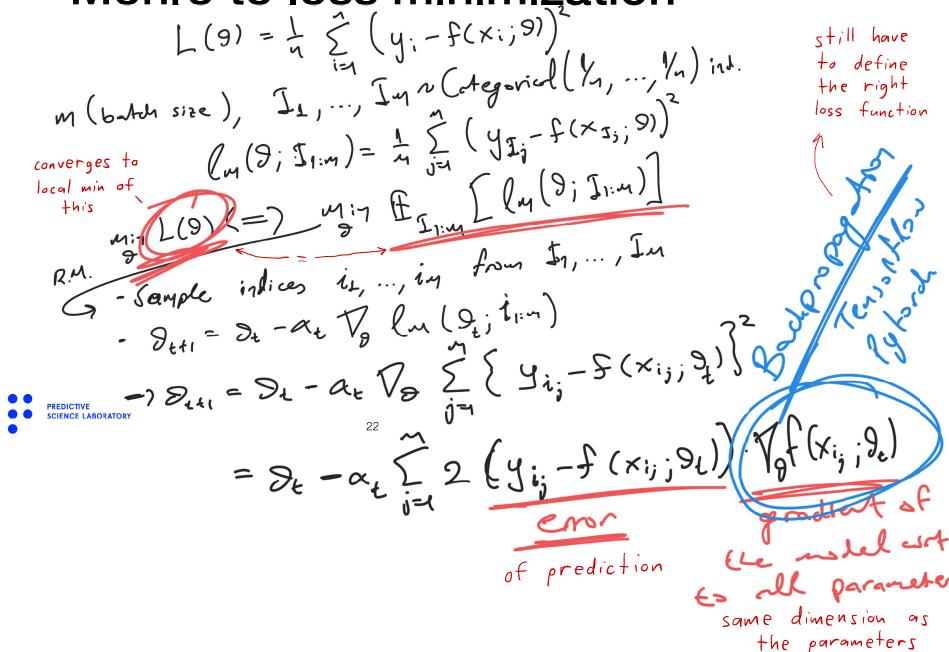
1. Initalize Do.

RM theorem the algorithm above will contege L a Geal minimum of the st. opt. (*) if $\sum_{t=1}^{NE} \alpha_t^2 < \alpha$ $\sum_{t=1}^{NE} \alpha_t^2 < \alpha$

Eat Alling
$$\alpha_t = +00$$
; $\alpha_t^2 < \alpha_t^2 < \alpha_t^2 < \alpha_t^2 < \alpha_t^2 = +00$; $\alpha_t^2 < \alpha_t^2 < \alpha_t^2$

$$\alpha_t = \frac{A}{(Bt+C)^2}, \quad 0.5 < e < 1$$

Application of Robbins-Monro to loss minimization



Advanced variants of stochastic gradient descent

Lato avoid having to select the learning rate

- Stochastic gradient descent with momentum.
- AdaGrad.
- Adam (adaptive moment estimation).

