**Foundations of Deep Learning – Homework Assignment #3**Adi Album & Tomer Epshtein

**Part 2: (2)**

Question: (Bonus)

Show that by decaying learning rate it is possible to guarantee convergence to expected -stationary point. In particular, define a scheme for and a respective upper bound on that will ensure:

Proof:

Let be a twice continuously differentiable and -smooth function that attains its global minimum  
.

We define an update scheme for   
In part(2) 1 we derived

Plugging in SGD’s update scheme where , we achieve:

We’ll analyze each separately.

* Because is independent of
* Because is independent of and , and from linearity of expected value with inner product, LHS is independent of .
* Because is independent of
* Because is independent of and , and from linearity of expected value with inner product, LHS is independent of .
* Because is independent of and

Bringing it all together we achieve

So

Assume that for steps we didn’t achieve an expected -stationary point.  
I.e. for all .

Define by . is convex monotonically increasing.  
So, for all , .

is convex, so by Jensen’s inequality:

I.e.

So, for steps

So:

This inequality holds for all

Since is ’s global minimum,   
So,

Where , so:

Side note:

Where did the update rule for , , come from?

Denote , In part 2(1) we showed:

We would like to select an update scheme for such that the above bound () is minimal.

Let efine .  
 is a parabola that attains its minimum at   
(This can easily be seen noticing is a parabola of the form with so its minimum is achieved at )

Selecting this update scheme gives us the desired bound