Menglu Liang
mjl6416@psu.edu

Homework

Spring PHS597
March 9, 2021

Derivation of the scaling factor for ADAM

We have (1)

$$S^{[k+1]} = \rho * S^{[k]} + (1-\rho)g \tag{2}$$

$$S^{[0]} = 0 \tag{3}$$

By (1) and (2), we then will have:

$S^{[1]} = (1-\rho)g$

$S^{[2]} = \rho * S^{[1]} + (1-\rho)g$

$= \rho * [(1-\rho)g] + (1-\rho)g$

$= (1+\rho)(1-\rho)g$:

$S^{[k+1]} = (1 + ... + \rho^k)(1-\rho)g$

$\Rightarrow \mathrm{E}(S^{[k+1]}) = \mathrm{E}((1-\rho^{k+1})g)$

$= (1-\rho^{k+1})\mathrm{E}(g)$