
Algorithm 1 AdaGrad

Require: Cost function $J(\mathbf{w})$, initial $\mathbf{w}^{(0)}$, initial $\mathbf{G}^{(0)}$, constants η and ϵ .

- 1: Set $i \leftarrow 0$
 - 2: **while** $\|\mathbf{w}^{(i)} - \mathbf{w}^{(i-1)}\| > \epsilon$ **do**
 - 3: Compute gradient vector $\mathbf{g}^{(i)} = \nabla J(\mathbf{w}^{(i)})$
 - 4: Accumulate $\mathbf{G}^{(i+1)} \leftarrow \mathbf{G}^{(i)} + \mathbf{g}^{(i)} \cdot \mathbf{g}^{(i)}$
 - 5: Update parameters $\mathbf{w}^{(i+1)} \leftarrow \mathbf{w}^{(i)} - \frac{\eta}{\sqrt{\mathbf{G}^{(i+1)} + \epsilon}} \mathbf{g}^{(i)}$
 - 6: Update $i \leftarrow i + 1$
 - 7: **end while**
 - 8: Return $\hat{\mathbf{w}} \leftarrow \mathbf{w}^{(i-1)}$
-