HW 6

IZ Raad

1/21/2024

1

What is the difference between gradient descent and *stochastic* gradient descent as discussed in class? (*You need not give full details of each algorithm. Instead you can describe what each does and provide the update step for each. Make sure that in providing the update step for each algorithm you emphasize what is different and why.*)

Student Input

2

Consider the `FedAve` algorithm. In its most compact form we said the update step is $\omega_{t+1} = \omega_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} \nabla F_k(\omega_t)$. However, we also emphasized a more intuitive, yet equivalent, formulation given by $\omega_{t+1}^k = \omega_t - \eta \nabla F_k(\omega_t); w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$.

Prove that these two formulations are equivalent.
(*Hint: show that if you place ω_{t+1}^k from the first equation (of the second formulation) into the second equation (of the second formulation), this second formulation will reduce to exactly the first formulation.*)

$$
\begin{aligned}
\omega_{t+1} &= \omega_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} \nabla F_k(\omega_t) \\
&= \omega_t - \sum_{k=1}^{K} \frac{n_k}{n} \eta \nabla F_k(\omega_t) \\
&= \omega_t - \sum_{k=1}^{K} \frac{n_k}{n} (\omega_t - \omega_{t+1}^k) \\
&= \sum_{k=1}^{K} \frac{\omega_t}{K} - \frac{n_k}{n} (\omega_t - \omega_{t+1}^k) \\
&= \sum_{k=1}^{K} (\frac{1}{K} - \frac{n_k}{n}) \omega_t + \frac{n_k}{n} \omega_{t+1}^k \\
&= \omega_t \sum_{k=1}^{K} (\frac{1}{K} - \frac{n_k}{n}) + \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k \\
&= \omega_t (1 - 1) + \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k \\
&= \sum_{k=1}^{K} \frac{n_k}{n} \omega_{t+1}^k
\end{aligned}
$$

3

Now give a brief explanation as to why the second formulation is more intuitive. That is, you should be able to explain broadly what this update is doing.

In our second formulation, we say that our data is split up into K partitions. Then, in each of these partitions, we locally progress one step of gradient descent. We then take the weighted average of each of these local steps of gradient descent (adjusted by partition size). This weighted average becomes our global step update.

4

Explain how the harm principle places a constraint on personal autonomy. Then, discuss whether the harm principle is *currently* applicable to machine learning models. (*Hint: recall our discussions in the moral philosophy primer as to what grounds agency. You should in effect be arguing whether ML models have achieved agency enough to limit the autonomy of the users of said algorithms.*)

Student Input