HW 6

Student Name

11/19/2024

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.*)

In both gradient descent and stochastic gradient descent, we update our solution in the opposite direction of the gradient with a learning rate α. The update rule for gradient descent is $\theta_{i+1} = \theta_i - \alpha \nabla f(\theta_i, X, Y)$. The difference between gradient descent and stochastic gradient descent is that stochastic gradient descent uses a random subset of the data to avoid falling into local minimums rather than global. As such, the update rule is $\theta_{i+1} = \theta_i - \alpha \nabla f(\theta_i, X_{i'}, Y_{i'})$ where $X_{i'}$ and $Y_{i'}$ are subsets of the original data.

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.*)

Since $\omega_{t+1}^k = \omega_t - \eta \nabla F_k(\omega_t)$, then $w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} (\omega_t - \eta \nabla F_k(\omega_t))$.
Since ω_t doesn't depend on k, we can rewrite as $w_{t+1} = \omega_t \sum_{k=1}^{K} \frac{n_k}{n} - \eta \sum_{k=1}^{K} \nabla F_k(\omega_t)$.
Here, $\sum_{k=1}^{K} \frac{n_k}{n} = 1$ so $w_{t+1} = \omega_t - \eta \sum_{k=1}^{K} \nabla F_k(\omega_t)$, thus they are equivalent.

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.

This formulation is more intuitive because it more closely resembles the general form for gradient descent, where we can see the gradient being subtracted from the previous omega. Then the second equation is where we take the weighted average.

Prove that randomized-response differential privacy is ϵ-differentially private.

Let A be a randomized algorithm. Consider a binary question where people report the truth a with probability p, or they flip their answer to 1-a with probability 1-p. Now consider D, D' to be datasets of responses that differ by one response.
Consider the case when the output is a.
Then $\frac{P[A(1)=1]}{P[A(0)=1]} = \frac{P[Output=1|Input=1]}{P[Output=1|Input=0]} = \frac{p}{1-p}$
Also, $\frac{P[A(0)=1]}{P[A(1)=1]} = \frac{1-p}{p}$
Thus $\frac{P[A(D)\in 1]}{P[A(D')\in 1]} = \frac{p}{1-p} = e^{ln(\frac{p}{1-p})}$

Define the harm principle. 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.*)

The harm principle essentially states that a person's free will is only truly free up until it harms someone else. I would argue that the harm principle is applicable to machine learning models. The main example that comes to my mind is the COMPAS algorithm. That is a case where an algorithm was subject to the harm principle, since it was not implemented due to concerns of its biases. If it had been implemented, it could have caused serious harm to people. Even if COMPAS had not been implemented as an end-all be-all decision maker and merely as an tool to assist decisions, it's a slipper slope that could set a bad precedent for other algorithms which have the ability to cause harm. Algorithms are becoming more prominent and clearly have the ability to impact people's decisions, so I would consider that enough agency to adhere to the harm principle.