HW 6

Ryan Dee

1/21/2024

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.*)

Student Input For Gradient Descent all data is used in finding the minimum or maximum, for stochastic gradient descent, all data is used for computing the gradient, but this can result in the method being "trapped" in a local minimum. However, Stochastic Gradient Descent uses a subset of the data to escape local minimum and arrive at the true minimum.

The update step for Gradient Descent is : $\theta_{(i+1)} = \theta_(i) - \alpha \nabla f(\theta_{(i)}, X_{(i)}, Y_{(i)}$

This is different from the Stochastic Gradient Descent update step: $\theta_{(i+1)} = \theta_i - \alpha \nabla f(\theta_i, X_i', Y_i')$ which uses a subset of the data, denoted Xi, and Yi, so that the method finds the true gradient descent, and doesn't get stuck at a local minimum #

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.)

Student Input

$\omega_{i+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

$\omega_{i+1} = \sum_{k=1}^{K} \frac{n_k}{n} (\omega_i - \alpha \nabla F_k(\omega_i))$

$\omega_{i+1} = \omega_i \sum_{k=1}^{K} \frac{n_k}{k} - \alpha \sum_{k=1}^{K} \frac{n_k}{k} \nabla F_k(\omega_i))$

$\omega_{i+1} = \omega_i - \alpha \sum_{k=1}^{K} (\frac{n_k}{k}) \nabla F_k(w_i))$

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.

Student Input This is more intuitive because you now get an update for each client and then average it out for a global update, instead of averaging out each client before updating.
# Prove that randomized-response differential privacy is ϵ-deferentially private.

Student Input A data set is said to be ϵ differential private if and only if for every data set D1, and D2 differing in exactly one element and subsets S is an element of the image of A : $\frac{P[A(D(1)\epsilon S]}{P[A(D2)\epsilon S]} <_e^\epsilon$

Randomized response is differential private where ϵ = ln3 PF: Observe that D and S ϵ {Yes, No} where S= Yes.

$\frac{P[A(Yes)=Yes]}{P[A(No)=Yes]} = \frac{P[Output=Yes|Input=Yes]}{P[Output=Yes|Input=No]} = \frac{3/4}{1/4} = e^{ln(3)}$

Then: $\frac{P[A(D(1)\epsilon S]}{P[A(D2)\epsilon S]} <= 3 = e^{(ln3)}$

This concludes the randomized response is differential private # 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.*)

Student Input Def(harm principle): Personal autonomy is checked at the point where exercising that autonomy causes objective harm to another agent.
Machine learning models do not rise to the level of autonomy that other moral agents do. In this case they are not protected from the harm principle. In the case of the harm principle there must be harm done. How can one harm a machine learning model? Does deleting the file the code is in harm the model? Does ripping up the paper that has the math of a model harm it? Today we would consider machine learning models property, COMPAS is property of Tim Brennan and David Wells, the same way my car is my property. If I were to crash my car intentionally, did I violate the harm principle for my car? I would say no, and I think the same could be said of the owners of COMPAS, did they violate the harm principle by deleting the algorithm? Most would say no, because how can you harm property that is unfeeling and unconscious.