HW 6

Sydney Mason

11/12/2024

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.*)

Gradient descent uses all available data to find the minimum of the dataset. Stochastic gradient descent uses a random part of the overall dataset with each step to reduce the risk of only finding a local minimum rather than the global minimum. The gradient is found by obtaining the partial derivatives of the dataset and making it negative in order to find the direction of steepest descent. The update step for stochastic gradient descent is $\theta_{i+1} = \theta_i - \alpha\nabla f(\theta_i, x_i, y_i)$ The update for gradient descent is $\theta_{i+1} = \theta_i - \alpha\nabla f(\theta_i, x, y)$ The x_i and y_i are the domains chosen for each individual step.

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.*)

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

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.

The second formula is more intuitive because it separates the two steps, the update of the local model and the averaging of the local updates. They are in fact equivalent, but the way that the second model is presented makes it clear that there are two steps to the overall process.

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

$\frac{\Pr[\mathcal{A}(\text{Yes})=\text{Yes}]}{\Pr[\mathcal{A}(\text{No})=\text{Yes}]} = \frac{\Pr[\text{Output=Yes}|\text{Input=Yes}]}{\Pr[\text{Output=Yes}|\text{Input=No}]} = \frac{p}{1-p} = e^{\ln\left(\frac{p}{1-p}\right)}$

$\frac{\Pr[\mathcal{A}(\text{No})=\text{Yes}]}{\Pr[\mathcal{A}(\text{Yes})=\text{Yes}]} = \frac{1-p}{p} = e^{\ln\left(\frac{1-p}{p}\right)}$

$p = \frac{e^\epsilon}{1+e^\epsilon}$ and $1-p = \frac{1}{1+e^\epsilon}$. The maximum is e^ϵ, satisfying differential privacy.

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 is designed so that a person's autonomy is given up until the point where it can harm another moral agent. Therefore, a person or entity has agency until they are seen to be harming someone. I don't believe that ML models have achieved enough agency to limit the autonomy of users of said algorithms. They are very much still learning, and are going off of a wide variety of information, information that might have conflicting ideas of what harm may be. They are unequipped to make decisions about the autonomy of the users. Also, a moral agent is defined to be someone who has autonomy and can be held accountable, and right now, ML models do have some autonomy, but cannot fully be held responsible for the outcomes of their decisions, meaning they are not in fact moral agents. Therefore, we cannot apply the harm principle to them.