HW7: Ethics and DP-SGD

CS 208 Applied Privacy for Data Science, Spring 2022

Version 1.0: Due Fri, Mar. 25, 5:00pm.

1. **Embedded EthiCS assignment:** See separate pdf.

2. **DP-SGD:** In our code example in class,[1] we saw how to release an estimated Logistic regression for predicting marital status from education level using DP-SGD to optimize the log-likelihood loss function. Convert this code to release the probability of employment given education level and disability status (these are the same variables we used in the Opacus example). You will need to modify the loss function, the gradient clipping, and Gaussian noise addition to handle the additional variable present here compared to the code from class.

   As discussed in the class, the learning rate parameter needs to be set correctly in order to obtain convergence in the DP-SGD setting. The learning rate ν in the notebook is a 2-dim vector (the coefficient of education level and the intercept).[2] Since we are now adding one more prediction variable, we need a third learning rate parameter, which you are going to find privately. (You can keep the education coefficient and intercept learning rates the same as in the notebook.)

   Run your code and create $K = 10$ differentially private models (each with privacy parameter $\epsilon = 1$ and $\delta = 1e - 6$) across a sequence of learning rates. (You can leave all the other parameters as in the exemplar code, or adjust them to reasonable values.) Choose one of these models to release, by means of the exponential mechanism with privacy parameter $\epsilon = 1$. Use a score function in the exponential mechanism that is the negative of the loss function you used in training. Show the parameters of the DP model that is trained using the chosen learning rate.

   The privacy loss of the entire procedure above can be analyzed by applying standard composition theorems, but this will incur a loss that grows with the number K of models trained. However, a theorem of Liu and Talwar (STOC 2019) shows that in fact one can do this model selection with only a constant-factor increase in the privacy-loss parameter ϵ, with no dependence on K.

3. **Revised Project Ideas:** You should have received feedback on your initial project ideas on Gradescope. Based on this feedback, we would like you to revise your ideas and create your project group (1-3 people per group). By **Monday 3/21**, please post your current topic ideas in the Google sheet we shared in Ed. With this homework submission on **Friday 3/25**, you should settle on your project group and submit your revised ideas for further feedback. Feel free to talk to any of the course staff in more detail as you are thinking through your revised idea. Please read the "Final Project Guidelines" (`https://github.com/opendp/cs208/blob/main/spring2022/final%20project/Final%20Project%20Guidelines.pdf`) document

[1] See `https://github.com/opendp/cs208/blob/main/spring2022/examples/wk6_dpsgd_full.ipynb`

[2] Ordinarily, the learning rate parameter is treated as a single scalar that multiplies the entire gradient. Allowing different learning rates per coordinate amounts to also normalizing the different independent variables.

on the course website and submit a paragraph as described in the “Topic Ideas (revision after Spring Break)” bullet.