STOR 390: HOMEWORK 7

Riley Harper

April 20, 2024

Abstract: This homework explores optimization methods in machine learning, specifically Gradient Descent (GD) and Stochastic Gradient Descent (SGD), and their fundamental differences, particularly in the context of the update step. Additionally, it delves into the FedAve algorithm, presenting a proof of equivalence between two of its formulations and offering an intuitive explanation of its federated learning approach. Lastly, it examines the harm principle's relevance to machine learning, questioning the agency of ML models and their potential to limit user autonomy through the unintended consequences of their application.

Question 1

Recall that in class we showed that for randomized response differential privacy based on a fair coin (that is a coin that lands heads up with probability 0.5), the estimated proportion of incriminating observations $\widehat{P}$ (in class this was the estimated proportion of students having actually cheated) was given by $\widehat{P} = 2\pi - \frac{1}{2}$ where π is the proportion of people answering affirmative to the incriminating question. I want you to generalize this result for a potentially biased coin. That is, for a differentially private mechanism that uses a coin landing heads up with probability $0 \leq \theta \leq 1$, find an estimate $\widehat{P}$ for the proportion of incriminating observations. This expression should be in terms of θ and π.

To generalize the result of the randomized response mechanism for a biased coin, we follow the methodology applied with a fair coin but adjust it to account for the bias in the coin flip probability.

Randomized Response Mechanism Setup:

1. **Options for Each Respondent:**
    - **Truthfully answer** the question (e.g., such as having not cheated).
    - **Lie** according to the outcome of a coin flip.

    For a biased coin:
    - Probability of flipping heads (θ) – the respondent lies.
    - Probability of flipping tails $(1 - \theta)$ – the respondent answers truthfully.
2. **If the coin comes up heads**, the respondent answers "Yes" or "No" irrespective of the truth and solely based on the result of the second coin flip.
3. **If the coin comes up tails**, the respondent answers truthfully.

Modeling Responses:

- Let π be the proportion of people who answer "Yes" to the question.
- When asked the question under this mechanism,
    - The probability of a respondent saying "Yes" because they are telling the truth (coin lands tails and they are in the truthful group): $(1 - \theta)\widehat{P}$.
    - The probability of a respondent saying "Yes" because they are lying (coin lands heads twice): θ^2.

So, the total probability of receiving a "Yes" response, denoted $\widehat{P}$, is given by the addition of the two outcomes: $\pi = \theta^2 + (1 - \theta)\widehat{P}$

Solving for π:

We want to express π in terms of P and θ. Rearranging the above equation:

$$\pi = \theta^2 + (1-\theta)\hat{P}$$
$$\implies \pi - \theta^2 = (1-\theta)\hat{P}$$
$$\implies \hat{P} = \frac{\pi-\theta^2}{1-\theta}$$

Question 2

Next, show that this expression reduces to our result from class in the special case where $\theta = \frac{1}{2}$.

$$\hat{P} = \frac{\pi - \theta^2}{1-\theta}$$

Substituting θ with $\frac{1}{2}$,

$$\hat{P} = \frac{\pi - (\frac{1}{2})^2}{1-(\frac{1}{2})}$$

Simplifying the numerator and denominator,

$$\hat{P} = \frac{\pi - \frac{1}{4}}{\frac{1}{2}}$$

Simplifying,

$$\hat{P} = 2\pi - \frac{1}{2}$$

This is the expression we derived in class for the estimated proportion of incriminating observations when using a fair coin for randomized response differential privacy.

Question 3

Consider the additive feature attribution model: $g(x') = \phi_0 + \sum_{i=1}^{M} \phi_i x_i'$ where we are aiming to explain prediction f with model g around input x with simplified input x'. Moreover, M is the number of input features.

Give an expression for the explanation model g in the case where all attributes are meaningless, and interpret this expression. Secondly, give an expression for the relative contribution of feature i to the explanation model.

Not covered so this was intentionally left blank.

Question 4

Part of having an explainable model is being able to implement the algorithm from scratch. Let's try and do this with KNN. Write a function entitled $chebychev$ that takes in two vectors and outputs the Chebychev or L^∞ distance between said vectors. I will test your function on two vectors below. Then, write a $nearest\_neighbors$ function that finds the user specified k nearest neighbors according to a user specified distance function (in this case L^∞) to a user specified data point observation.

Chebyshev distance: 6

KNN Data:

54	78	25
77	87	31
88	34	82
80	100	95
58	84	79
96	4	98
17	48	75
100	51	31
47	80	82
11	94	36
25	54	78
31	77	87
82	88	34
95	80	100
79	58	84
98	96	4
75	17	48
31	100	51
82	47	80
36	11	94

Test Point:

23	23	23

Nearest Neighbors:

17	48	75
75	17	48
54	78	25

Question 5

Finally create a *knn_classifier* function that takes the nearest neighbors specified from the above functions and assigns a class label based on the mode class label within these nearest neighbors. I will then test your functions by finding the five nearest neighbors to the very last observation in the $iris$ dataset according to the $chebychev$ distance and classifying this function accordingly.

Nearest Neighbors to the Observation:

6	5.4	3.9	1.7	0.4
5	5.0	3.6	1.4	0.2
6.1	5.4	3.9	1.7	0.4
5.1	5.0	3.6	1.4	0.2
6.2	5.4	3.9	1.7	0.4

Features of the Observation to be Classified:

Sepal.Length	5.9
Sepal.Width	3.0
Petal.Length	5.1
Petal.Width	1.8

Predicted Class of the Observation: setosa

Actual Class of the Observation: virginica

Question 6

Interpret this output. Did you get the correct classification? Also, if you specified $K = 5$, why do you have 7 observations included in the output dataframe?

The predicted class label by KNN was setosa, however, the true class was that of virginica. The code I wrote for KNN did not include what was written in class to include 7 observations, however, if I had written it in the same manner these additional outputs would have been the distances associated with the neighbor list.

Question 7

Earlier in this unit we learned about Google's DeepMind assisting in the management of acute kidney injury. Assistance in the health care sector is always welcome, particularly if it benefits the well-being of the patient. Even so, algorithmic assistance necessitates the acquisition and retention of sensitive health care data. With this in mind, who should be privy to this sensitive information? In particular, is data transfer allowed if the company managing the software is subsumed? Should the data be made available to insurance companies who could use this to better calibrate their actuarial risk but also deny care? Stake a position and defend it using principles discussed from the class.

Sensitive health care data exists in a critical, complex intersection between technological innovation and personal privacy. The use of artificial intelligence, as exemplified by Google's DeepMind in managing acute kidney injury, illustrates the potential of technology to enhance patient care. However, patient privacy is a cornerstone of the healthcare system, underpinning the principles of autonomy and consent. When patients share their information with healthcare providers, there is an implicit understanding that this exchange is for the sole purpose of enhancing their health and well-being. The moment this information becomes a commodity that can be transferred, especially without explicit consent, this foundational trust is violated.

In the context of corporate acquisition, where the company responsible for managing healthcare software is absorbed by another entity, the continuity of data privacy policies becomes uncertain. It is imperative that patient data is not merely bundled into the assets transferred during such corporate movements. Instead, there must be a deliberate, transparent process that ensures patients are informed and their consent is obtained anew, thereby respecting their autonomy over personal information. When it comes to sharing patient data with insurance companies, the stakes are considerably higher. Insurance providers could argue that access to comprehensive data enables them to accurately calibrate actuarial risks. However, there is a thin line between informed risk calculation and the potential for discriminatory practices. If sensitive health data were freely available to these companies, it could lead to a situation where patients are penalized for conditions beyond their control, facing denial of coverage or prohibitive premiums. This not only challenges the ethical principle of justice, which advocates for fairness and equity, but also the principle of nonmaleficence, obliging us to refrain from causing harm to others. Furthermore, the governance of such data should be meticulously regulated to prevent misuse. Legislation akin to the Health Insurance Portability and Accountability Act (HIPAA) offers a framework that emphasizes patient rights and privacy. Strong regulatory oversight can ensure that insurance companies, should they be granted access, are not allowed to use this information to unjustly discriminate against individuals in need of coverage.

In advocating for stringent privacy controls and limited access to sensitive health care data, the principles of transparency and accountability cannot be overemphasized. Patients deserve clarity regarding the use of their data, with robust mechanisms in place to hold entities accountable for their data practices. Opting out of data sharing that does not directly contribute to personal health care should always be an option available to patients.

Overall, while the integration of AI into healthcare is to be welcomed for its ability to enhance patient care, it must not be pursued at the expense of compromising patient privacy and the ethical standards that govern the healthcare profession. The priority must always be the rights and well-being of the individual patient, ensuring that advancements in technology are matched with equally strong protections for sensitive health information.