Differential Privacy for Regularised Linear Regression (slides)

Ashish Dandekar, Debabrota Basu, Stéphane Bressan

July 18, 2019
A randomised algorithm $\mathcal{M}$ with domain $\mathcal{D}$ is $\epsilon$-differentially private if for all $S \in \text{Range}(\mathcal{M})$ and $D, D' \in \mathcal{D}$ such that $D$ and $D'$ are neighbouring datasets

$$\log \left( \left| \frac{\Pr(\mathcal{M}(D) \in S)}{\Pr(\mathcal{M}(D') \in S)} \right| \right) \leq \epsilon$$

Privacy preserving mechanisms that satisfy the definition of $\epsilon$-differential privacy are $\epsilon$-differentially private mechanisms.
Linear regression uses a linear function to map predictor attributes, $x_i \in \mathbb{R}^d$, of a data point $t_i = (x_i, y_i)$ to its response attribute, $y_i \in \mathbb{R}$.
Linear Regression
The parameters $\theta \in \mathbb{R}^d$ are estimated by minimising mean squared loss over the training dataset, $T$.

$$\theta^* = \arg \min_{\theta} \left( \frac{1}{|T|} \sum_{t_i \in T} (x_i^t \theta - y_i)^2 \right)$$
Objective function for regularised linear regressions

**Ridge**

$$\theta^* = \arg \min_{\theta} \left( \left( \frac{1}{n} (X\theta - Y)^2 \right) + \lambda \|\theta\|_2^2 \right)$$

**LASSO**

$$\theta^* = \arg \min_{\theta} \left( \left( \frac{1}{n} (X\theta - Y)^2 \right) + \lambda \|\theta\|_1 \right)$$

**Elastic net**

$$\theta^* = \arg \min_{\theta} \left( \left( \frac{1}{n} (X\theta - Y)^2 \right) + \lambda (\alpha \|\theta\|_2^2 + (1 - \alpha) \|\theta\|_1) \right)$$
The functional mechanism is a privacy preserving mechanism that introduces noise in the objective function that is minimised to estimate the parameters of a machine learning model.

Zhang et al. propose a functional mechanism. [?] that satisfies $\epsilon$-differential privacy. It does so by adding a calibrated amount of Laplace noise in the coefficients of the Taylor series expansion of the function of the linear regression.
Satisfying $\epsilon$-differential privacy

**Sensitivity**

Sensitivity of a function $f$, denoted as $\Delta_f$, is the maximum change in the output when any pair of neighbouring datasets is given to the function.

$$\Delta_f = \max_{x \sim y} \| f(x) - f(y) \|_1$$

**Laplace Mechanism**

Dwork et al. [?] propose a privacy preserving mechanism that adds Laplace random noise to the output of a function, $f$. If the scale of the Laplace random variable is set to $\frac{\Delta_f}{\epsilon}$, the mechanism satisfies $\epsilon$-differential privacy.
Satisfying $\epsilon$-differential privacy

Sensitivity of the mean squared loss

In [?], Zhang et al. show that the sensitivity of the loss function less than to two times the sum of maximum values of coefficients in the Taylor expansion.

Taylor expansion of the mean squared loss

$$\text{loss}_T(\theta) = \frac{1}{|T|} \sum_{t_i \in T} (x_i^t \theta - y_i)^2$$

$$= \frac{1}{|T|} \sum_{(x_i, y_i) \in T} (y_i)^2 - \frac{1}{|T|} \sum_{j=1}^{d} \left( 2 \sum_{d_i \in T} y_i x_{ij} \right) \theta_j$$

$$+ \frac{1}{|T|} \sum_{1 \leq j, l \leq d} \left( \sum_{d_i \in T} x_{ij} x_{jl} \right) \theta_j \theta_l$$
Satisfying $\epsilon$-differential privacy

If we normalize all the attributes in any dataset, $D$, of size $n$ such that they lie in $[-1, 1]$,

$$\Delta_{loss}(\theta) \leq \frac{2}{n} \max_{t_i \in D} \left( y^2 + 2 \sum_{j=1}^{d} yx_{ij} + \sum_{1 \leq j, l \leq d} x_{ij} x_{jl} \right)$$

$$= \frac{2}{n} \left( 1 + 2d + d^2 \right)$$
Summing up...

For a training dataset \( T \) with \( n \) data points, the predictor attributes can be represented as a matrix \( X \in \mathbb{R}^{n \times d} \) and the response attribute as a vector \( Y \in \mathbb{R}^n \). In this notation, the loss function is written as:

\[
\text{loss}_T(\theta) = \frac{1}{n} \left( \theta^t (X^t X) \theta - 2 \theta^t (X^t Y) + Y^t Y \right)
\]

\[
= \frac{1}{n} \left( \theta^t (M) \theta - 2 \theta^t (N) + O \right)
\]

Laplace mechanism is used that adds Laplace\( (0, \frac{\Delta f}{\epsilon}) \) noise in \( M, N \) and \( O \). And the output of the functional mechanism, \( \theta^* \), is calculated as:

\[
\theta^* = \arg \min_{\theta} \frac{1}{n} \left( \theta^t (M_{\text{noisy}}) \theta - 2 \theta^t (N_{\text{noisy}}) + O_{\text{noisy}} \right)
\]
Canonical Census Dataset
We conduct experiments on a subset of the 2000 US census dataset provided by Minnesota Population Center in its Integrated Public Use Microdata Series [?]. We consider 316,276 records of the heads of households in our dataset.

Each record has 6 attributes, namely, Age, Gender, Race, Marital Status, Education, Income out of which Age and Income are numerical attributes and rest of them are categorical attributes. Regression analysis is performed using Income as the response attribute and the rest of the attributes as predictor attributes.
Boxplot of RMSE of elastic net regression with functional mechanism for different values of $\epsilon$ for the census dataset.
Results

- **Linear Regression**
  - RMSE vs $\varepsilon$
  - "functional LR" vs "linear LR"

- **Ridge Regression**
  - RMSE vs $\varepsilon$
  - "functional RG" vs "RG" vs "LR"

- **LASSO Regression**
  - RMSE vs $\varepsilon$
  - "functional LS" vs "LS" vs "LR"

- **Elastic net regression**
  - RMSE vs $\varepsilon$
  - "functional EN" vs "EN" vs "LR"
Discussion

Instability in the result

Addition of noise to the quadratic term in the Taylor expansion leads to loss in the convexity of the objective function. Non-convex objective functions do not possess a global minimum. Hence, the local optima give rise to unstable results.

Validity of the privacy guarantee

In [], functional mechanism proves the privacy guarantee of the loss function. Authors say that the same proof naturally leads to the privacy guarantee of the parameters of the regression.
Conclusion