CoinPress Extension: Linear Regression

ediger.t

March 2021

1 Problem

The problem we are trying to address is to privately solve linear regression. If we assume we have a distribution $(\vec{x}, y)$, with $\vec{x} \sim \mathcal{N}(\mu, \Sigma)$ and $y|\vec{x} \sim \mathcal{N}(\langle \vec{x}, \beta \rangle, \sigma)$, then the closed form solution to approximate $\hat{\beta}$ is

$$\hat{\beta} = (X^T X)^{-1} X^T y = (\frac{1}{n} X^T X)^{-1} (\frac{1}{n} X^T y)$$

Where $\frac{1}{n} X^T X$ converges to the inverse of the covariance matrix of X, and $\frac{1}{n} X^T y$ converges to the expected value $E(x * \langle x, \beta \rangle)$. By turning this into a problem of estimating mean and covariance, we can use CoinPress on the two, and combine their results to find an private estimate for $\hat{\beta}$.

2 Extension with Simplifying Assumptions

If we first consider the case where $\vec{x} \sim N(0, I_{dxd})$, and $y|\vec{x} \sim N(\langle \vec{x}, \beta \rangle, \sigma^2)$. Then we know $\frac{1}{n} X^T X = I_{dxd}$, so

$$\hat{\beta} = I \frac{1}{n} X^T y = \frac{1}{n} X^T y$$

Therefore, we need to estimate $\frac{1}{n} X^T y \approx E(x * \langle x, \beta \rangle)$. We can do this by using CoinPress mean estimation with with input $z_i = x_i y_i$.

2.1 Getting our extension to work with CoinPress

To estimate $\hat{\beta}$ we are using the CoinPress estimators for the mean on $z_i = x_i y_i$, for $i \in \{1, ..., n\}$. The the algorithm that estimates multivariate mean is MVMRec. MVMRec takes in n samples $X_{1...n}$ from $N(\mu, I_{d \times d})$, $B_2(c, r)$ containing μ, $t \in N^+$, $\rho_{1...n}, \beta > 0$.

2.1.1 Rescaling Covariance

MVMRec is stated and implemented as an algorithm for a Gaussian with identity variance, but the same argument works for an arbitrary known covariance

Σ if we rescale the data. The expected covariance of the z_i's is in fact not the identity so a rescale is necessary. In particular, we can analyze what the actual covariance matrix of the z_i's would be by looking at what the values for the diagonals, and off-diagonals of the covariance matrix.

** INSERT MATH done to find diagonal, non-diagonal entries of cov **

We find that the diagonal entries ($Cov(z_j, z_j)$) take on the value $\beta_j^2 + \|\beta\|_2^2 + 1$, and the off-diagonal entries ($Cov(z_j, z_k)$) take on the value $\beta_j \beta_k$. Therefore $\Sigma \in R^{d\times d}$, with $\Sigma = \beta\beta^T + (\|\beta\|_2^2 + 1)I_{d\times d}$.

INSERT PICTURE/ REASON why we can use $2\|\beta\|_2^2 + 1$ to approx Σ

Therefore, in order to use MVMRec, we will need to privately estimate norm of β.

2.1.2 How to define B_2

2.1.3 Choosing number of iterations t

2.1.4 Choosing $\rho_{1\ldots n}$

2.2 Privacy

2.3 Results