Note: This problem set will be applied. You are asked to submit specific results, but please attach your code to the back of your entire submission (i.e. don't put code in the solution area for every problem, just staple it to the back).

Pretend you are working for the Washington D.C. as a data analyst collaborating with the "Capital Bikeshare" bikesharing program. Initially, they took bikes away for servicing while people were trying to use the bikes, which drove away customers. Your boss asked you to find the 'best' way to take bikes away for routine maintenance in that the fewest number of customers are looking for bikes when they are being serviced. She hands you a somewhat large ($n \sim 18,000$ points and $p = 36$ features) matrix $\boldsymbol{A} \in \mathbb{R}^{n \times p}$ where every row represents an hour in the bike-share system and every column represents a feature like the weather during that hour, as well as the corresponding vector $\boldsymbol{b} \in \mathbb{R}^n$ of outputs, which are the log-number of users in the bike-share system. See the raw data at `github.com/math389L/math389L.github.io/raw/master/data/bikes.csv` as well as a `README` file at `github.com/math389L/math389L.github.io/raw/master/data/bikes.readme.txt`. A sample file in Python which processes the raw data and creates the matrices $\boldsymbol{A}$ and $\boldsymbol{b}$ is found at `github.com/math389L/math389L.github.io/raw/master/hw/pset3_loading_data.py`. Your task will be to solve the linear regression problem $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ for $\boldsymbol{x}$ in a number of ways and compare them.

---

1 Solve the least squares problem $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ for the minimizer $\boldsymbol{x}^\star$ exactly using a built-in method. We will refer to this as $\boldsymbol{x}^\star$ for the rest of this problem set. Repeating this a number of times, report the average time it takes to compute $\boldsymbol{x}^\star$ using the built-in method. Answer the following question by looking at $\boldsymbol{x}^\star$: do people in Washington D.C. tend to use the bikeshare program less when the wind speed increases?

■

---

2 Solve the least squares problem $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ for the minimizer $\boldsymbol{x}_{\mathsf{chol}}$ by using a Cholesky factorization to solve the normal equations $\boldsymbol{A}^* \boldsymbol{Ax} = \boldsymbol{A}^* \boldsymbol{b}$ explicitly. Report the relative error
$$\frac{\|\boldsymbol{Ax}_{\mathsf{chol}} - \boldsymbol{b}\|_2}{\|\boldsymbol{Ax}^\star - \boldsymbol{b}\|_2}$$
of your solution using this method instead of a built-in solver.

■

3 Solve the least squares problem $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ for the minimizer $\boldsymbol{x}_{\mathrm{cw}}$ by first applying a Clarkson-Woodruff sketch $\boldsymbol{S}$ to compute $\boldsymbol{SA}$ and $\boldsymbol{Sb}$ before finally using a built-in least squares solver to compute $\boldsymbol{x}_{\mathrm{cw}} = \arg\min \|\boldsymbol{SAx} - \boldsymbol{Sb}\|_2$. Plot a histogram of 1000 samples of the relative error (approximation ratio)

$$\frac{\|\boldsymbol{Ax}_{\mathrm{cw}} - \boldsymbol{b}\|_2}{\|\boldsymbol{Ax}^\star - \boldsymbol{b}\|_2}$$

using this method for $k = 50$. Repeat this for $k = 75, 100, 250$. For $k = 250$ report the average time it takes on your computer to compute the sketches and solve for $\boldsymbol{x}_{\mathrm{cw}}$ and compare this to the time to compute the solution using a built-in solver.

■

4 Solve the least squares problem $\min \|\boldsymbol{Ax} - \boldsymbol{b}\|_2$ for the minimizer $\boldsymbol{x}_{\mathrm{sgd}}$ computed using stochastic gradient descent as we analyzed in Problem 2 of the second problem set. Plot a histogram of 1000 runs of the relative error (approximation ratio)

$$\frac{\|\boldsymbol{Ax}_{\mathrm{sgd}} - \boldsymbol{b}\|_2}{\|\boldsymbol{Ax}^\star - \boldsymbol{b}\|_2}$$

using this method for a number of steps $t = 50$. Repeat this for $t = 500, 1000, 2500$. For $t = 1000$ report the average time it takes on your computer to compute $\boldsymbol{x}_{\mathrm{sgd}}$ and compare this to the time to compute the solution using a built-in solver, as well as the sketching method used in Problem 3.

■