CSE 546 Homework 2B

Jake Gonzales

November 10, 2023

Collaborators: n/a

1 Convexity and Norms

B1.

(a) To show that $f(x) = \|x\|$ is a convex function we will use the definition of convexity. Let $\lambda \in [0, 1]$ and apply the triangle inequality and absolute scalability for any $x, y \in \mathbb{R}^n$:

$$\|\lambda x + (1 - \lambda)y\| \leq \|\lambda x\| + \|(1 - \lambda)y\| \leq \lambda\|x\| + (1 - \lambda)\|y\|.$$

Therefore we have shown $f(x) = \|x\|$ is a convex function.

(b) We want to show that the set $D := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is a convex set. Let $\lambda \in [0, 1]$ then from the previous problem in B1-a we have that

$$\|\lambda x\| + \|(1 - \lambda)y\| \leq \lambda\|x\| + (1 - \lambda)\|y\|.$$

From the defined set we say that $\|x\| \leq 1$ and $\|y\| \leq 1$ for any $x, y \in D$. Then we can further write

$$\|\lambda x\| + \|(1 - \lambda)y\| \leq \lambda\|x\| + (1 - \lambda)\|y\| \leq \lambda + 1 - \lambda = 1.$$
$$\implies \lambda x + (1 - \lambda)y \in D.$$

Thus we have shown that any line segment between the two points lies in D hence it is a convex set.

(c) The set $D := \{(x, y) : g(x, y) \leq 4\}$ where $g(x, y) = (|x|^{1/2} + |y|^{1/2})^2$ is below. This set is not convex, it is clear to see we can take a point at one of of the "spikey points" of the set and try to draw a line to the other point and it would leave the set, thus breaking the definition of convexity.

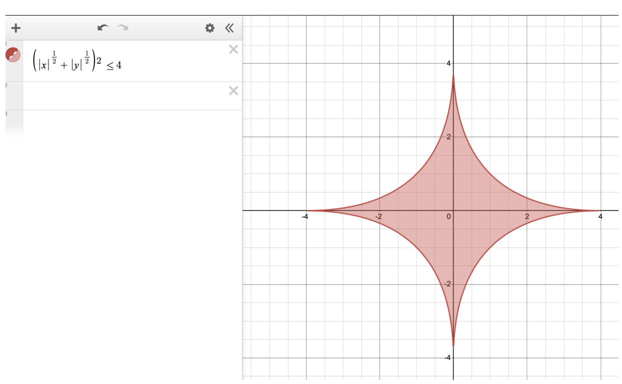

Figure 1: The set D defined by g(x,y) in the problem statement.

2 Bounding the Estimate

B2.

(a) By definition we can write the least squares estimator as follows:

$$\begin{aligned}
\widehat{\beta} &= (X^TX)^{-1}X^TY \\
&= (X^TX)^{-1}X^T(X\beta^* + \epsilon) \\
&= \beta^* + (X^TX)^{-1}X^T\epsilon \\
\implies \widehat{\beta} - \beta^* &= (X^TX)^{-1}X^T\epsilon \\
\widehat{\beta} &= \beta^* + (X^TX)^{-1}X^T\epsilon
\end{aligned}$$

Since $\widehat{\beta}$ is normally distributed and $\epsilon \sim \mathcal{N}(0, I)$ and assuming X is fixed:

$$\begin{aligned}
\widehat{\beta} &\sim \mathcal{N}(0 * (X^TX)^{-1}X^T\epsilon + \beta^*, (X^TX)^{-1}X^TI((X^TX)^{-1}X^T)^T) \\
&= \mathcal{N}(\beta^*, (X^TX)^{-1}X^T((X^TX)^{-1}X^T)^T) \\
&= \mathcal{N}(\beta^*, (X^TX)^{-1}X^TX(X^{-1}X^T)^T) \\
&= \mathcal{N}(\beta^*, X^{-1}X^{-T}) \\
&= \mathcal{N}(\beta^*, (X^TX)^{-1})
\end{aligned}$$

where $(X^TX)^{-1}$ is the covariance matrix of $\widehat{\beta}$. For each of the elements $\widehat{\beta}_j$ the variance will be the $j-th$ diagonal element of the covariacne matrix. And, since $\widehat{\beta}_j$ is normally distributed, all of its elements will be as well> Thus we write:

$$\widehat{\beta}_j = \mathcal{N}(B_j^*(X^TX)^{-1}_{i,j}$$

(b) Let's fix $\delta \in (0, 1)$ and let $j \in [d]$. From the previous problem we can derive

$$\begin{aligned}
\widehat{\beta}_j &= \mathcal{N}(B_j^*(X^TX)^{-1}_{i,j} \\
\implies \frac{\widehat{\beta}_j}{(X^TX)^{-1}_{j,j}} &\sim \mathcal{N}(0, 1)
\end{aligned}$$

Then by proposition 2, we have that with a probability of at least $1 - \delta$, $\widehat{\beta}_j$ will be contained within a confidence interval:

$$\begin{aligned}
&\left[\beta^* - \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}, \quad \beta^* + \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}\right] \\
&= \left[-\sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}, \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}\right]
\end{aligned}$$

Therefore,

$$\mathbb{P}\left(-\sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)} \leq \widehat{\beta}_j \leq \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}\right) \geq 1 - \delta$$

$$\implies \mathcal{P}\left(\left|\widehat{\beta}_j\right| \leq \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}\right) \geq 1 - \delta.$$

We <u>cannot</u> conclude that with probability at least $1-\delta, \left|\widehat{\beta}_j\right| \leq \sqrt{2(X^TX)^{-1}_{j,j}\log(2/\delta)}$ for all $j \in [d]$ simultaneously because each of the events have a probability of at least $1-\delta$ but this does not guarantee that all of the events with simultaneously be at least probability of $1-\delta$.

(c) If we run the code below we see that we have 63 out of 10,000 $\widehat{\beta}_j's$ outside of the interval.

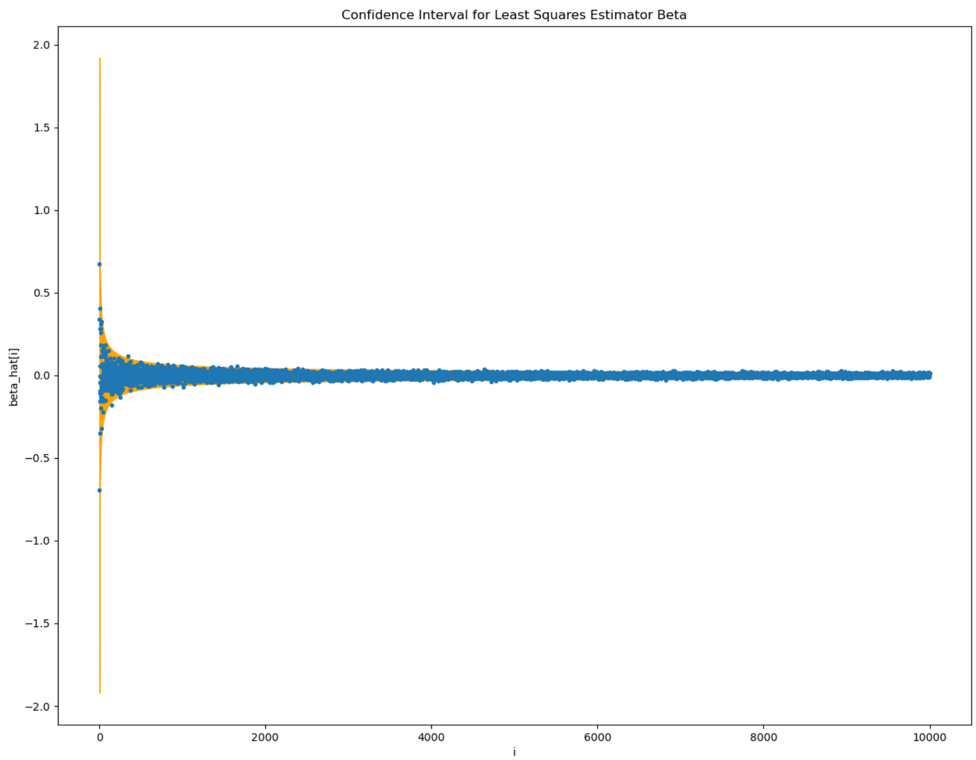

Figure 2: Confidence Interval

```
import numpy as np
from matplotlib import pyplot as plt
import random
random.seed(10)
delta = 0.05
n = 20000
d = 10000
beta = np.zeros(d)
```

```python
X = np.empty((n,d))
inv_XTX = np.zeros((d,d))
y = np.random.normal(0, 1, n)
for i in range(n):
    index = (i%d)+1
    e = np.zeros(d)
    e[index-1] = 1.0
    X[i,:] = np.sqrt(index)*e

XTX = X.T @ X
ci = np.zeros(d)

for i in range(d):
    inv_XTX[i,i] = 1/(XTX[i,i])
    ci[i] =np.sqrt(2*inv_XTX[i,i]*np.log(2/delta))

beta_hat = inv_XTX @ X.T @ y

count = 0
for i in range(d):
    if np.abs(beta_hat[i])> ci[i]:
        count += 1
x = np.linspace(1,d,d)
plt.figure(figsize=(20,20))
plt.plot(x,beta_hat, '.')
plt.fill_between(x, (beta-ci), (beta+ci), color='orange')
plt.xlabel("i")
plt.ylabel("beta_hat[i]")
plt.title("Confidence Interval for Least Squares Estimator Beta")
plt.show()
print(count)
```