APPM 5360, Spring 2023 - Homework 1 and 2

Eappen Nelluvelil

February 3, 2023

1. Show that a set is convex if and only if its intersection with any line is convex. Show that a set is affine if and only if its intersection with any line is affine.

   (a) ($\Longrightarrow$) Let $C \subseteq \mathbb{R}^n$ be a given convex set, i.e., for all $x, y \in C$ and for all $t \in [0,1]$, $tx + (1-t)\,y \in C$. Also, let z be a given line in $\mathbb{R}^n$, i.e., $z = \theta x_1 + (1-\theta)\,x_2$, for some $x_1, x_2 \in \mathbb{R}^n$, with $x_1 \neq x_2$ and $\theta \in \mathbb{R}$.

   We now consider $C \cap z$. Let $x, y \in C \cap z$ be given, with $x \neq y$, and let $t \in [0,1]$ be given. Necessarily, $tx + (1-t)\,y$ lies on z since it is a line, and $tx + (1-t)\,y \in C$ since C is convex. Thus, $C \cap z$ is convex.

   ($\Longleftarrow$) Let $C \subset \mathbb{R}^n$ be a given set such that its intersection with any line $z = \theta x_1 + (1-\theta)\,x_2$, where $x_1, x_2 \in \mathbb{R}^n$, $x_1 \neq x_2$, and $\theta \in \mathbb{R}$, is convex.

   Let $x, y \in C$ be given, with $x \neq y$, and $t \in [0,1]$ be given. Consider $tx + (1-t)\,y$. If we consider the line that contains x and y, then necessarily, $tx + (1-t)\,y$ lies on the line, and the intersection of C with this line will contain $tx + (1-t)\,y$ since the intersection is convex. Since this holds for any $x, y \in C$, C is convex.

   (b) ($\Longrightarrow$) Let $C \subseteq \mathbb{R}^n$ be a given affine set, i.e., for all $x, y \in C$ and for all $t \in \mathbb{R}$, $tx + (1-t)\,y \in C$. Also, let z be a given line in $\mathbb{R}^n$, i.e., $z = \theta x_1 + (1-\theta)\,x_2$, for some $x_1, x_2 \in \mathbb{R}^n$, with $x_1 \neq x_2$ and $\theta \in \mathbb{R}$.

   We now consider $C \cap z$. Let $x, y \in C \cap z$ be given, with $x \neq y$, and let $t \in \mathbb{R}$ be given. Necessarily, $tx + (1-t)\,y$ lies on z since it is a line, and $tx + (1-t)\,y \in C$ since C is affine. Thus, $C \cap z$ is affine.

   ($\Longleftarrow$) Let $C \subset \mathbb{R}^n$ be a given set such that its intersection with any line $z = \theta x_1 + (1-\theta)\,x_2$, where $x_1, x_2 \in \mathbb{R}^n$, $x_1 \neq x_2$, and $\theta \in \mathbb{R}$, is affine.

   Let $x, y \in C$ be given, with $x \neq y$, and $t \in \mathbb{R}$ be given. Consider $tx + (1-t)\,y$. If we consider the line that contains x and y, then necessarily, $tx + (1-t)\,y$ lies on the line, and the intersection of C with this line will contain $tx + (1-t)\,y$ since the intersection is affine. Since this holds for any $x, y \in C$, C is affine.

2. Which of the following sets S are polyhedra? If possible, expression S in the form $S = \{x \mid Ax \preceq b, Fx = g\}$.

   (a) $S = \{y_1 a_1 + y_2 a_2 \mid -1 \leq y_1 \leq 1, -1 \leq y_2 \leq 1\}$, where $a_1, a_2 \in \mathbb{R}^n$.

This is a polyhedron, and is given by the convex hull of the vectors $a_1 + a_2$, $a_1 - a_2$, $-a_1 + a_2$, and $-a_1 - a_2$. Without loss of generality, we will analyze one face of this polyhedron, i.e., we consider the face that contains the direction vector from $-a_1 + a_2$ to $-a_1 - a_2$, which we will denote b_1, and the contains the direction vector from $a_1 + a_2$ to $-a_1 + a_2$, which we will denote b_2.
The nullspace of the matrix whose columns are b_1 and b_2 consists of vectors that are orthogonal to the plane spanned by b_1 and b_2, and we will denote one such element of this nullspace n_1. Thus, for x that lie in this face, we have that $n_1^T x = c_1$, where $c_1 = n_1^T x_0$, where x_0 is some vector that lies in the plane spanned by b_1 and b_2.
This gives one such half-space, and we can find a vector n_2 and a corresponding scalars c_2 such that $n_2^T x = c_2$. A normal vector for the side opposite to b_1 is given by $-n_1$, and a normal vector for the side opposite to b_2 is given by $-n_2$.
For an x that lies in the interior of the face, which is a parallelogram spanned by b_1 and b_2, we require x to satisfy the inequalities

$$
\begin{aligned}
n_1^T x &\le n_1^T b_2 \\
-n_1^T x &\le -n_1^T b_2, \\
n_2^T x &\le n_2^T b_1, \\
-n_2^T x &\le -n_2^T b_1.
\end{aligned}
$$

We can follow this procedure for the other faces of the polyhedron.

(b) $S = \left\{x \in \mathbb{R}^n \mid x \succeq 0, 1^T x = 1, \sum_{i=1}^n x_i a_i = b_1, \sum_{i=1}^n x_i a_i^2 = b_2\right\}$, where $a_1, \ldots, a_n \in \mathbb{R}$ and $b_1, b_2 \in \mathbb{R}$.

This is a polyhedron, as it is a finite intersection of halfspaces and hyerplanes. We first define $\widetilde{a}_1 = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T$ and $\widetilde{a}_2 = \begin{bmatrix} a_1^2 & a_2^2 & \cdots & a_n^2 \end{bmatrix}^T$. Then, the polyhedron can be defined by

$$
\begin{aligned}
1^T x &\ge 0, \\
1^T x &= 1, \\
\widetilde{a}_1^T x &= b_1, \\
\widetilde{a}_2^T x &= b_2.
\end{aligned}
$$

(c) $S = \left\{x \in \mathbb{R}^n \mid x \succeq 0, x^T y \le 1 \text{for all } y \text{ with } ||y||_2 = 1\right\}$

This is not a polyhedron as it is the intersection of the non-negative orthant (in n dimesions) with the unit ball in $\mathbb{R}^n$, which cannot be described in a finite number of intersections of half-spaces and hyerplanes.

(d) $S = \left\{x \in \mathbb{R}^n \mid x \succeq 0, x^T y \le 1 \text{for all } y \text{ with } \sum_{i=1}^n |y_i| = 1\right\}$

This is a polyhedron. Since $x \succeq 0$ and $x^T y \le 1$ for all y such that $||y||_1 = 1$, this implies that $||x||_1 \le 1$. Thus, we can describe all such x that satisfy the above conditions via the following inequalities:

$$
\begin{aligned}
-x &\preceq 0, \\
1^T x &\le 1,
\end{aligned}
$$

which means this set is the finite intersection of half-spaces and hyerplanes.

3. Which of the following sets are convex?

(a) A *slab*, i.e., a set of the form $S = \left\{x \in \mathbb{R}^n \mid \alpha \le a^T x, x \le \beta\right\}$.

Let $x, y \in S$ be given, and let $t \in [0,1]$ be given. We have that $\alpha \leq a^T x \leq \beta$, and $\alpha \leq a^T y \leq \beta$. We also have that

$$a^T \left(tx + (1-t)\, y\right) = a^T x + (1-t)\, a^T y.$$

We have that

$$\begin{aligned} \alpha &= t\alpha + (1-t)\,\alpha \\ &\leq ta^T x + (1-t)\, a^T y \end{aligned}$$

and

$$\begin{aligned} ta^T x + (1-t)\, a^T y &\leq t\beta + (1-t)\,\beta \\ &= \beta. \end{aligned}$$

Thus, a slab is convex.

(b) A *rectangle*, i.e., a set of the form $S = \{x \in \mathbb{R}^n \mid \alpha_i \leq x_i \leq \beta_i, i = 1, \ldots, n\}$. A rectangle is sometimes called a hyperrectangle when $n > 2$.

Let $x, y \in S$ be given, and let $t \in [0,1]$ be given. Define $\alpha = \begin{bmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_n\end{bmatrix}$ and $\beta = \begin{bmatrix}\beta_1 & \beta_2 & \cdots & \beta_n\end{bmatrix}$. We know that $\alpha \preceq x \preceq \beta$ and $\alpha \preceq y \preceq \beta$. Consider $tx + (1-t)\, y$. Then

$$\begin{aligned} \alpha &= t\alpha + (1-t)\,\alpha \\ &\preceq tx + (1-t)\, y \\ &\preceq t\beta + (1-t)\,\beta \\ &= \beta. \end{aligned}$$

Thus, a rectangle is convex.

(c) A *wedge*, i.e., $S = \left\{x \in \mathbb{R}^n \mid a_1^T x \leq b_1, a_2^T x \leq b_2\right\}$.

Let $s, y \in S$ be given, and let $t \in [0,1]$ be given. We have that

$$\begin{aligned} a_1^T x &\leq b_1, \\ a_2^T x &\leq b_2, \\ a_1^T y &\leq b_1, \\ a_2^T y &\leq b_2. \end{aligned}$$

Then,

$$\begin{aligned} a_1^T \left(tx + (1-t)\, y\right) &= ta_1^T + (1-t)\, a_1^T y \\ &\leq tb_1 + (1-t)\, b_1 \\ &= b_1 \end{aligned}$$

and

$$\begin{aligned} a_2^T \left(tx + (1-t)\, y\right) &= ta_2^T x + (1-t)\, a_2^T y \\ &= tb_2 + (1-t)\, b_2 \\ &= b_2. \end{aligned}$$

Thus, a wedge is convex.

(d) The set of points closer to a given point than a given set, i.e.,

$$R = \{x \mid ||x - x_0||_2 \leq ||x - y||_2 \text{ for all } y \in S\},$$

where $S \subseteq \mathbb{R}^n$.

For a fixed point $y \in S$, the set of all points that are closer to x_0 than y is a half-space, which is convex. Then, R is the arbitrary intersection of half-spaces, and since the arbitrary intersection of convex sets is convex, R is convex.

(e) The set of points closer to one set than another, i.e.,

$$R = \{x \mid \text{dist}(x, S) \leq \text{dist}(x, T)\},$$

where $S, T \subseteq \mathbb{R}^n$, and

$$\text{dist}(x, S) = \inf \{||x - z||_2 \mid z \in S\}.$$

The set R is not convex. Without loss of generality, we will work in $\mathbb{R}^2$. Consider a set $S \subset \mathbb{R}^2$, which consists of two disjoint closed balls of the same radius $r_1 > 0$ and centers x and y, which we will denote $B_r(x)$ and $B_r(y)$, respectively. Let T be a closed ball with radius $r_2 > 0$ and center z such that it is disjoint from S and z lies on the line containing x and y and is equidistant from x and y.
If we consider a point x_1 that is closer to $B_r(x)$ than T and x_2 that is closer to $B_r(y)$ than T, then we can form a convex combination of x_1 and x_2 that lies outside of S, e.g. $\frac{1}{2}x_1 + \frac{1}{2}x_2$.
Thus, the set is not convex.

(f) The set $R = \{x \mid x + S_2 \subseteq S_1\}$, where $S_1, S_2 \subset \mathbb{R}^n$, with S_1 convex.

Let $y \in S_2$ be given, and let $x_1, x_2 \in R$ be given. This implies that

$$x_1 + y \in S_1,$$
$$x_2 + y \in S_1.$$

Let $t \in [0, 1]$ be given. Then, we consider the point $tx_1 + (1-t)x_2 + y$. Note that $t(x_1 + y) + (1-t)(x_2 + y) \in S_1$ since S_1 is convex. Expanding this sum, we have that

$$S_1 \ni t(x_1 + y) + (1-t)(x_2 + y) = tx_1 + (1-t)x_2 + y,$$

which implies that $tx_1 + (1-t)x_2 + y \in S_1$.
Since this holds for all points $y \in S_2$, we have that R is the arbitrary intersection of convex sets. Thus, we have that R is convex.

(g) The set of points whose distance to a does not exceed a fixed fraction θ of the distance to b, i.e., the set $S = \{x \mid ||x - a||_2 \leq \theta ||x - b||_2\}$. You can assume $a \neq b$ and $0 \leq \theta \leq 1$.

To show that this set S is convex, we will show that it is a ball, which is a convex set. Let $x \in S$ be given. We have that

$$\begin{aligned}
||x-a||_2 \leq \theta\,||x-b||_2 &\implies ||x-a||_2^2 \leq \theta^2\,||x-b||_2^2 \\
&\implies x^Tx - 2a^Tx + a^Ta \leq \theta^2\left(x^Tx - 2b^Tx + b^Tb\right) \\
&\implies \left(1-\theta^2\right)x^Tx - 2\left(a^T - \theta^2 b^T\right)x + a^Ta - \theta^2 b^Tb \leq 0 \\
&\implies x^Tx - \frac{2\left(a^Ta - \theta^2 b^Tb\right)}{1-\theta^2}x \leq \frac{\theta^2 b^Tb - a^Ta}{1-\theta^2}
\end{aligned}$$

We then complete the square on the term on the left by adding and subtracting $\left(\frac{a^Ta-\theta^2 b^Tb}{1-\theta^2}\right)^2$ to the left. This results in the following inequality:

$$\left|\left|x - \frac{a^Ta - \theta^2 b^Tb}{1-\theta^2}\right|\right|_2^2 \leq \left(\frac{\theta^2 b^Tb - a^Ta}{1-\theta^2} + \left(\frac{a^Ta - \theta^2 b^Tb}{1-\theta^2}\right)^2\right).$$

We see that this is the equation for a closed ball in $\mathbb{R}^n$, which is a convex set. Thus, the set is convex.

4. Consider the set of rank-k *outer products*, defined as $R = \left\{XX^T \mid X \in \mathbb{R}^{n\times k}, \text{rank}\,(X) = k\right\}$. Describe its conic hull in simple terms.

Note: the *conic hull* of a set C is the set of all conic combinations of points in C, and is also the smallest convex cone containing C. It is *not* just the smallest (possibly non-convex) cone containing C.

The set R consists of rank-k symmetric, positive semi-definite matrices that are of size $n \times n$. Without loss of generality, we will work with one element of R, which we denote XX^T. We have that XX^T is of rank-k and symmetric, positive semi-definite.

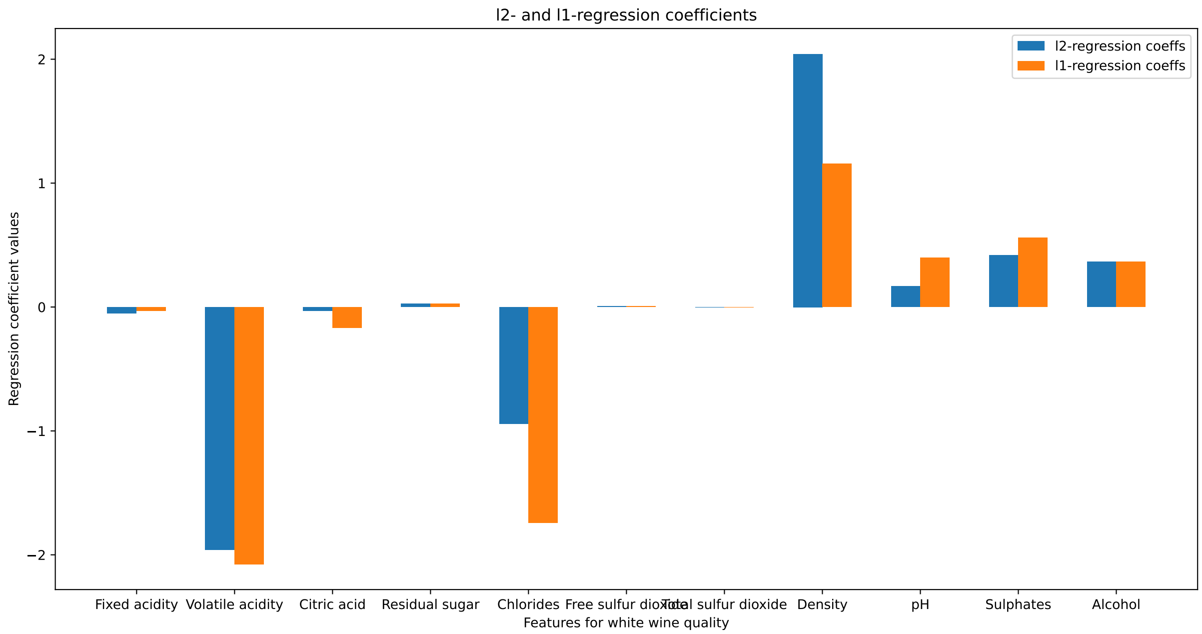
First, note that taking the non-negative sum of two rank-k, symmetric, positive semi-definite matrices does not decrease the rank of the resulting sum. For example, consider the non-negative sum of two rank-k, symmetric positive semi-definite matrices $X_1X_1^T$ and $X_2X_2^T$, i.e., $\theta_1X_1X_1^T + \theta_2X_2X_2^T$, where $\theta_1, \theta_2 \geq 0$. If x is in the nullspace of the above sum, we have that

$$\begin{aligned}
\left(\theta_1X_1X_1^T + \theta_2X_2X_2^T\right)x = 0 &\iff x^T\left(\theta_1X_1X_1^T + \theta_2X_2X_2^T\right)x = 0 \\
&\iff \theta_1x^TX_1X_1^Tx + \theta_2x^TX_2X_2^Tx = 0.
\end{aligned}$$

Since $X_1X_1^T$ and $X_2X_2^T$ are symmetric, positive semi-definite, the above implication means that $x^TX_1X_1^Tx = 0$ and $x^TX_2X_2^Tx = 0$. Thus, the dimension of the nullspace of the non-negative sum of rank-k, symmetric, positive semi-definite matrices cannot be greater than $n-k$, which implies that the rank of such a sum can only stay fixed at k or be greater by the rank-nullity theorem.

Thus, a conic combination of the elements of R is at a least rank-k symmetric, positive semi-definite matrix, with the rank possibly increasing. This means that the conic hull of R consists of all symmetric, positive semi-definite matrices that are between rank-k and rank-n, inclusive.

I solved first the two regression problems on the full data set using `cvxpy`, and I obtained the following regression coefficients, which I plotted on a bar plot:

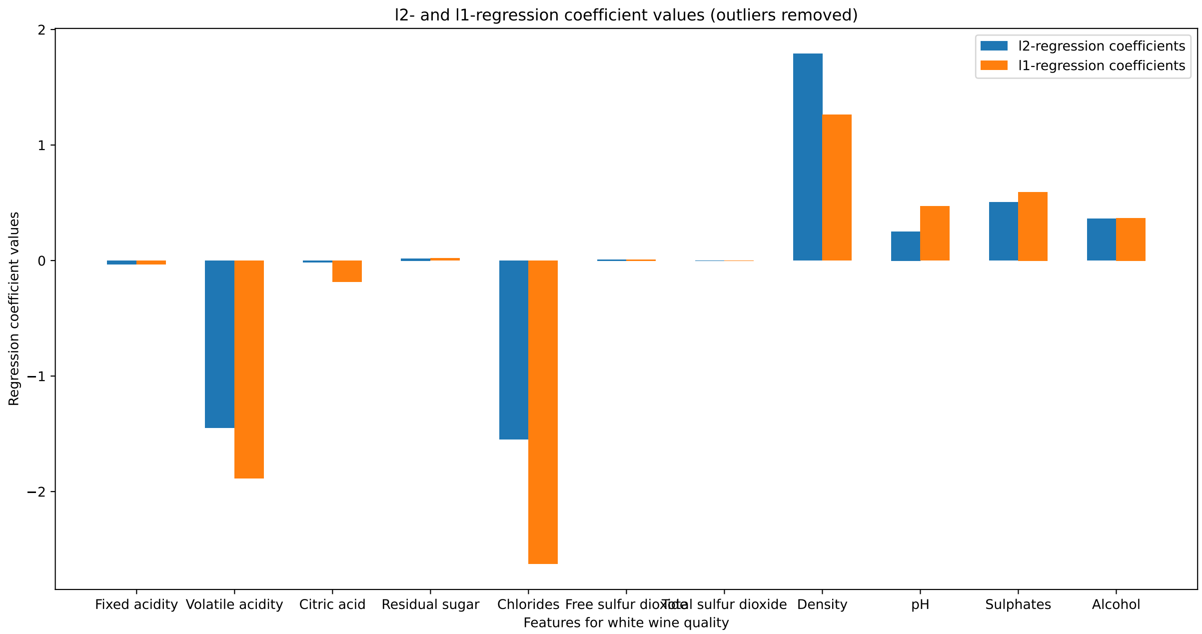


When comparing the values of the regression coefficients, feature-by-feature, the largest differences appeared in the "Chlorides" and "Density" features, with the values for the other features being comparable. Furthermore, `cvxpy` returned an optimal objective value of ≈ 52.82 for the least-squares regression problem and returned an optimal objective value of ≈ 2863 for the ℓ^1-regression problem. The higher objective value for the ℓ^1-regression problem is higher than the least squares problem due to the ℓ^1 problem being more susceptible to outliers.

To bring down the objective values, I removed outlier rows from the white wine data set. The procedure I followed was to compute the infinity norm of each row of the white wine data set and store the results in a vector. Then, I computed the interquartile range of the values in this vector, and used Boolean indexing to determine the rows whose infinity norms were below the 25^{th} and 75^{th} percentile of the computed infinity norms. After determining the outlier rows in this manner, I removed the outlier rows from the data set and recomputed the least-squares regression coefficients and ℓ^1 regression coefficients again. I plotted the new coefficients on a bar plot, which is shown below:

Removing the outliers in this manner lowered the objective values for both regression problems, but we see that the coefficients are not much more in agreement with one another. One idea I have to possibly address this issues is to normalize the entries in each feature column first before removing outlier rows, as computing the infinity norm of each row of the data set without prior manipulations might not be statistically sound.

The code used to solve all 4 problems and generate the above bar plots are given below.



Listing 1: Code used to solve regression problems and make bar plots

```
import numpy as np
import cvxpy as cvx
import matplotlib.pyplot as plt

white_wine_dataset = np.loadtxt("winequality-white.csv", delimiter=";",skiprows=1)
white_wine_features = ("Fixed acidity", "Volatile acidity", "Citric acid",
                       "Residual sugar", "Chlorides", "Free sulfur dioxide",
                       "Total sulfur dioxide", "Density", "pH",
                       "Sulphates", "Alcohol")

white_wine_quality = white_wine_dataset[:, -1]
white_wine_dataset = white_wine_dataset[:, 0:-1]

# Solve the the least-squares regression problem
beta1 = cvx.Variable(np.shape(white_wine_dataset)[1])
obj1 = cvx.Minimize(cvx.norm(white_wine_dataset @ beta1 - white_wine_quality, 2))
prob1 = cvx.Problem(obj1)
prob1.solve(verbose=False)

print("Optimal value for l2-regression:  ", prob1.value)

# Solve the l1-regression problem
beta2 = cvx.Variable(np.shape(white_wine_dataset[1]))
obj2 = cvx.Minimize(cvx.norm(white_wine_dataset @ beta2 - white_wine_quality, 1))
```

```
prob2 = cvx.Problem(obj2)
prob2.solve(verbose=False)

print("Problem value for l1-regression:  ", prob2.value)

# Compute the infinity-norm of the difference between the l2-regression
# coefficients and the l1-regression coefficients
inf_norm_regression_coeffs = np.linalg.norm(beta1.value - beta2.value, ord=np.inf)
print("Infinity norm between l2- and l1-regression coefficients: ", inf_norm_regression_co

# Make a bar plot of the l1- and l2-regression coefficients
width = 0.3
labels = np.arange(1, np.shape(white_wine_dataset)[1] + 1)

plt.figure(1)
plt.bar(labels, beta1.value, width, label="l2-regression coeffs")
plt.bar(labels + width, beta2.value, width, label="l1-regression coeffs")

plt.xticks(labels + width/2, white_wine_features)
plt.xlabel("Features for white wine quality")
plt.ylabel("Regression coefficient values")
plt.title("l2- and l1-regression coefficients")
plt.legend(loc="best")

# We remove the outliers by taking the infinity norm of each row of the dataset
# and using the IQR to remove rows whose infinity norms are below the 25th and
# are above the 75th percentiles
inf_norms_white_wine_dataset = np.linalg.norm(white_wine_dataset, ord=np.inf, axis=1)
num_rows = np.size(inf_norms_white_wine_dataset)
lower_quartile = np.median(np.sort(inf_norms_white_wine_dataset)[0:int(num_rows/2)])
upper_quartile = np.median(np.sort(inf_norms_white_wine_dataset)[int(num_rows/2):])

non_outlier_rows = np.logical_and(inf_norms_white_wine_dataset > lower_quartile,
                                  inf_norms_white_wine_dataset < upper_quartile)



white_wine_dataset_no_outliers = white_wine_dataset[non_outlier_rows, :]
white_wine_quality_no_outliers = white_wine_quality[non_outlier_rows]

# Solve the least-squares regression problem with outliers removed
beta3 = cvx.Variable(np.shape(white_wine_dataset_no_outliers)[1])
obj3 = cvx.Minimize(cvx.norm(white_wine_dataset_no_outliers @ beta3 - white_wine_quality_n
prob3 = cvx.Problem(obj3)
prob3.solve(verbose=False)

print("Optimal value for l2-regression  (outliers removed): ", prob3.value)

# Solve the l1-regression problem with outliers removed
beta4 = cvx.Variable(np.shape(white_wine_dataset_no_outliers)[1])
obj4 = cvx.Minimize(cvx.norm(white_wine_dataset_no_outliers @ beta4 - white_wine_quality_n
```

```python
prob4 = cvx.Problem(obj4)
prob4.solve(verbose=False)

print("Problem status for l1-regression (outliers removed): ", prob4.status)
print("Optimal value for l1-regression  (outliers removed): ", prob4.value)

# Make a bar plot of the l1- and l2- regression coefficients, after removing
# the outlier rows
plt.figure(2)
plt.bar(labels, beta3.value, width, label="l2-regression coefficients")
plt.bar(labels + width, beta4.value, width, label="l1-regression coefficients")

plt.xticks(labels + width/2, white_wine_features)
plt.xlabel("Features for white wine quality")
plt.ylabel("Regression coefficient values")
plt.title("l2- and l1-regression coefficient values (outliers removed)")
plt.legend(loc="best")
plt.show()
```

Listing 2: Output from the above code

```
Optimal value for l2-regression:   52.86155040697632
Problem value for l1-regression:   2863.0000933910037
Infinity norm between l2- and l1-regression coefficients:  0.8851261838815767
C:\Users\eappe\miniconda3\envs\cvxpy_env\lib\site-packages\cvxpy\problems\problem.py:1385:
  warnings.warn(
Optimal value for l2-regression  (outliers removed):  36.06178856441259
Problem status for l1-regression (outliers removed):  optimal
Optimal value for l1-regression  (outliers removed):  1372.427523194257
```