# STAT0028 Week 3 Exercises

Andrew Wu

2024-10-19

## Question 1

### (a)

$$\rho_{L_1}(u) = |u| = \begin{cases} u & u \geq 0 \\ -u & u < 0 \end{cases}$$

### (b)

$$\psi_{L_1}(u) = \begin{cases} 1 & u > 0 \\ -1 & u < 0 \end{cases}$$

We can extend the definition of $\psi$ to $u = 0$ with $\psi(0) = 0$, so that we have $\psi \equiv \operatorname{sgn}$. While the absolute value function is not differentiable at 0, it is a convex function and therefore has a subderivative (of 0) at the origin.

### (c)

The likelihood function is

$$\begin{aligned} \mathcal{L}(y, X|\beta) &\propto \prod_{i=1}^{n} e^{-\lambda|y_i - x_i^\top \beta|} \\ &= \exp\left(-\lambda \sum_{i=1}^{n} |y_i - x_i^\top \beta|\right). \end{aligned}$$

Since $\lambda > 0$, maximising the likelihood function is equivalent to minimising

$$\sum_{i=1}^{n} |y_i - x_i^\top \beta| = \sigma \sum_{i=1}^{n} \rho_{L_1}\left(|y_i - x_i^\top \beta|/\sigma\right),$$

which is exactly the objective function in L1 estimation of a general linear model.

## Question 2

### (a)

$$w(u) = \frac{\operatorname{sgn}(u)}{u}$$

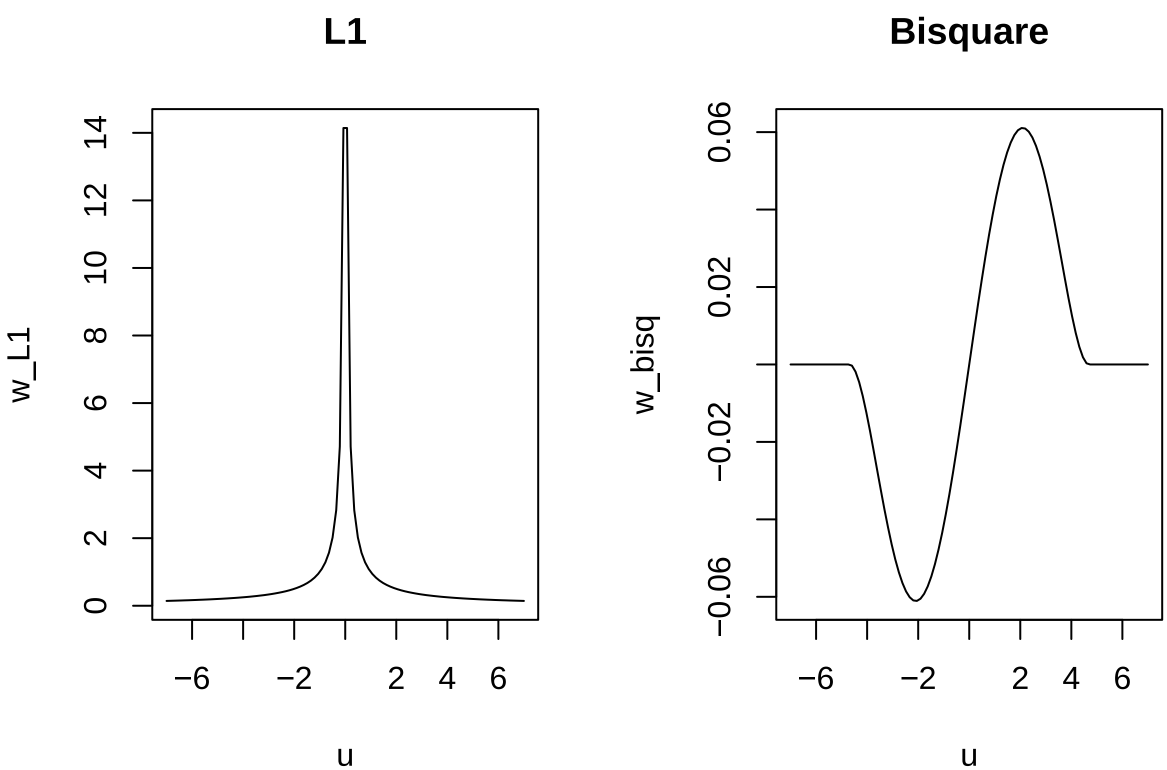
### (b)

$$w(u) = \begin{cases} \frac{u}{c^2}\left(1 - \frac{u^2}{c^2}\right)^2 & |u| \leq c \\ 0 & |u| > c. \end{cases}$$

```r
N = 100
umin = -7
umax = 7
u <- seq(umin, umax, length.out=N)
c <- 4.685

w_L1 <- sign(u) / u
w_bisq <- ifelse(abs(u) <= c, u/c^2 * (1 - u^2/c^2)^2, 0)

par(mfrow = c(1, 2))
plot(u, w_L1, type='l', main="L1")
plot(u, w_bisq, type='l', main="Bisquare")
```




- The L1 estimator only yields positive weights, while the Bisquare yields symmetrically positive or negative weights depending on the sign of the residual (below the threshold $c$). Negative weights are undesirable and are in practice often floored at 0.
- The L1 estimator has less robust weights than the Bisquare. For very small residuals, the L1 weight is unbounded, (whereas the bisquare approaches 0). For very moderately large residuals, the weight is non-zero, while it is exactly zero beyond $c$ for the bisquare. The L1 weights are also discontinuous at 0.

## Question 3

### (a)

The sign for all the coefficients are the same and magnitudes are similar though slightly larger in the robust models. This suggests all models identify the same general relation between the predictors and the response, which is reassuring. Unlike the LS model, the MM model yields significant (at a 1% level) p-values for all predictors including `x1`, likely due to ignoring the outlier in observation 19.

**(b)**

The least squares analysis identifies observations 10, 15 and 19 as outliers based on the Cook's distance and normality quantile analyses. In the MM model, only 19 is identified as an outlier and thus given zero weight, however, observations 10/15 also have relatively different weights compared to the rest of the data (as well as observations 8 and 20).

**(c)**

Both the LMS and MM models appear to perform worse than the OLS estimator when inspecting the plot of residuals vs fitted values. Both LMS and MM appear to violate homoskedasticity, with a clear downward trend in residuals for as fitted values increase. This suggests that both robust regression models are systematically underestimating the turnip vitamin B2 content for at larger values. Furthermore, normality seems to be more heavily violated based on the qqplot for MM compared with OLS. However, this is not so important in the Robust Regression setting - for example, we might even prefer alternative error distribution assumptions (consider q1, where the L1 estimator is also the MLE if the distribution is double exponential).