Exersize 4

(e) Discuss briefly on the results

Let's see the results of our fit on the data

-> LS estimator: Fairly good fit, with weights above $\theta_3$ ~~[illegible]~~ fluctuating between positive and negative values at large values. This is a sign of overfit meaning our model is heavily unconfident about new data and presents great variance as an estimator.

-> Ridge Estimator: Here we maintain a fairly good fit for values $10^{-4}$, $10^{-3}$, $10^{-2}$ and we see a great penalization on the terms above the $2^{nd}$ degree. Theta (3 +) values are relatively small compared to the lesser terms. However at $l_2 = 0,1$ we see that the penalization interferes with the quality of the fit, ledding to worse results.

→ **Lasso Estimation**: In our lasso regression estimation we see a good fit at $\ell_1 = 5 \cdot 10^{-4}$ and $\ell_1 = 10^{-4}$ with a strict penalization of terms other than $\theta_0$ and $\theta_2$, a hard assumption that our model is of $y = \theta_2 x_2^2 + \theta_0$ form. At values greater than that our model seems to over-penalize the fit, leading from a poor-fitted model to a complete 0-variance straight line at $\ell_1 = 0,1$.

# Exersize 6

(ii) Comment briefly on the results.

In exersize 5 we proved by example that the fit of a LSE on a random dataset, after some trials approaches the true values, under of course the assumption of a proper fit. That means our $\hat{\theta}$ estimator eventually converges to its dataset-generating value at high N (a lot of trials). Some fits may estimate higher, some lower but their mean "cancels-out" the deviation between them, leading to a better estimate. Also we proved that $l_2$ penalization (Ridge Reg.) is not always a solution. We need to pick a sensible value of $\ell_2$ in order to get good results. As I plotted in my Jupyter Notebook, a relatively low range of such values (2-40) gives the best results.