Q3 part a :

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is challenging as a value too small may result in a long training process that could get stuck, whereas a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process.

When we chose larger learning rate , we can achieve the desired regression in fewer iterations.

Q4 part a:

Yes. Because outliers have the same weight as others.

Q6 part a:

Normalization is performed on data to remove amplitude variation and only focus on the underlying distribution shape.

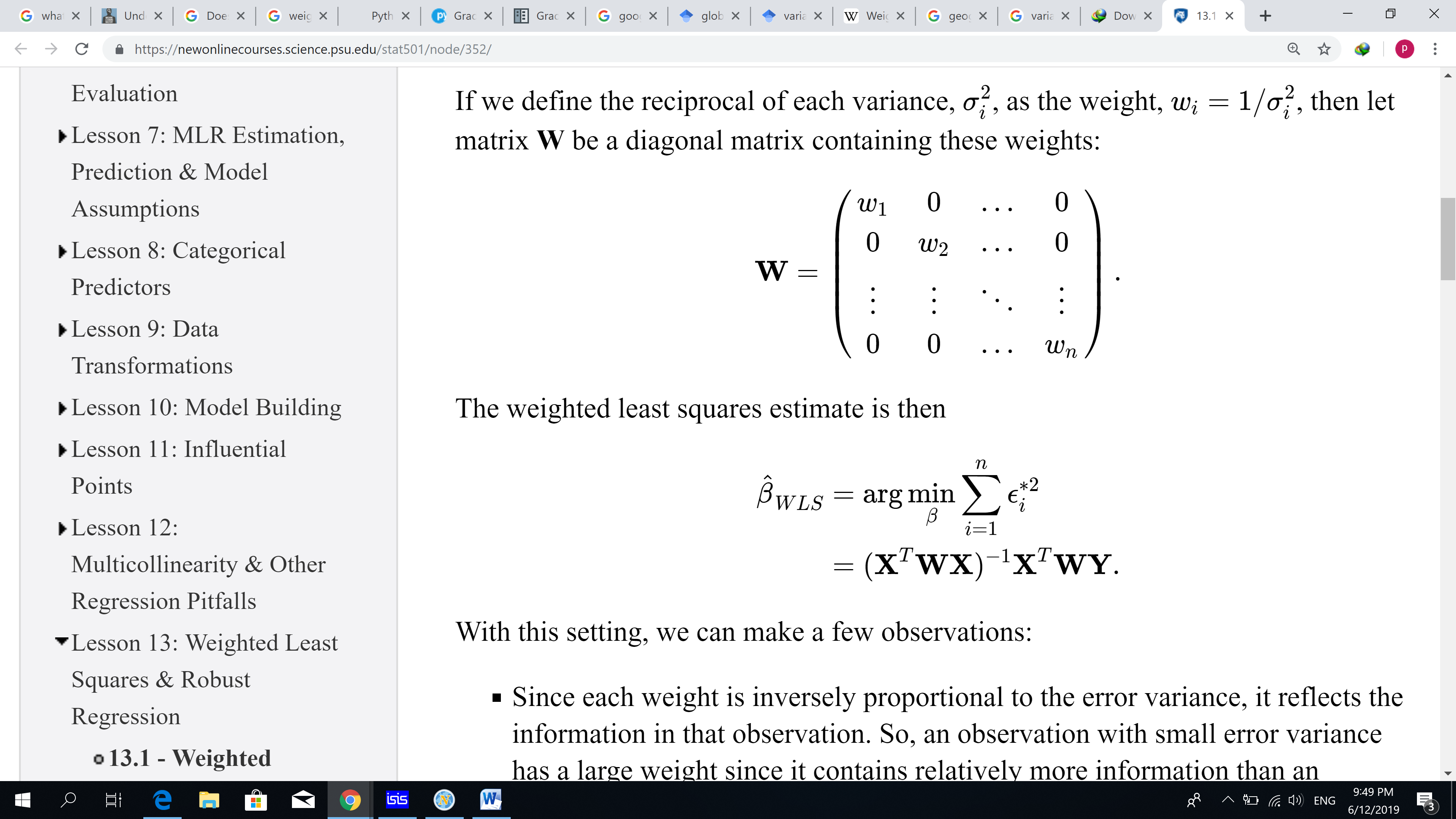
Normalization is performed on data to compare numeric values obtained from different scales.

Q2 part b:

We used a weighting function witch calculated the variance of each data i and defines

W(i) = 1/variance(i)

And created this weighting matrix:



this weighting function decrease the effect of datas witch are far from the center of the all data. Because the 1/variance of them is lower than others.

This method increase the robustness of the regression so this is appropriate. The regression is more dependant on realistic data

Q 6 part 2:

This method increase the robustness of the regression so this is appropriate. The regression is more dependant on realistic data