Below are the results for each of the three sections of the assignment. Early stopping helped in the regression case. As it is shown in the plots, validation loss initially goes up and then it decreases until it reaches its minimum. Then it starts to go up again. Early stopping helped to stop the algorithm when the validation loss reached its minimum. If early stopping is not used, the validation loss will increase progressively as the model starts to overfit.

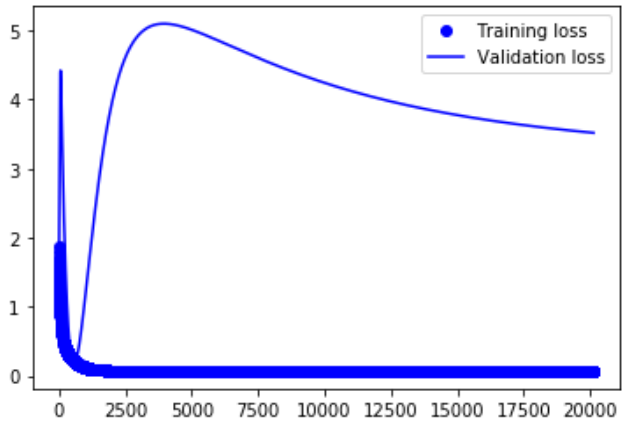
In the two cases of classification, early stopping did not help as the results would be the same with and without early stopping. In these cases, lower lambdas produced lower training and validation losses as it seemed the model did not overfit. Therefore, it seemed that higher regularization slightly impacted the model performance negatively.

With a decently chosen learning rate, running the algorithms seemed to take too long on my personal laptop, so I used 10^{-3} for the regression and 10^{-2} for the classification cases as the cut-off points for the gradient norm.

<u>Regression Under Quadratic Loss:</u>

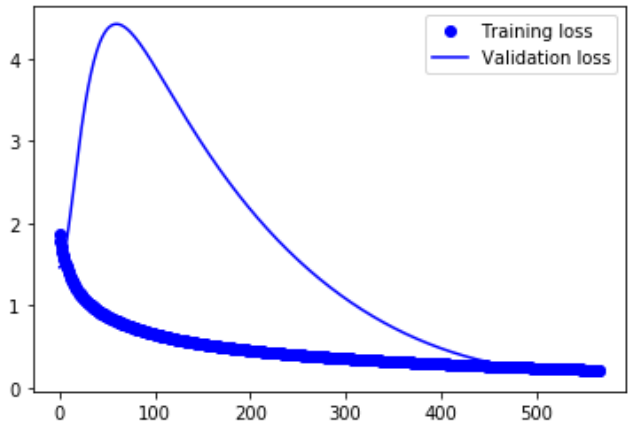
No Early Stopping:

λ	Train Error	Validation Error
0	0.06	4.45
0.0001	0.06	4.45
0.01	0.06	3.51



With Early Stopping:

λ	Train Error	Validation Error
0	0.21	0.17
0.0001	0.21	0.17
0.01	0.21	0.17



Classification Under Logistic Loss:

λ	Train Error	Validation Error	Train Acc	Validation Acc	Test Acc
0	0.12	0.12	98.67	98.18	98.13
0.0001	0.12	0.12	98.67	98.18	98.13
0.01	0.18	0.19	98.74	98.18	98.23

Classification Under Hinge Loss:

λ	Train Error	Validation Error	Train Acc	Validation Acc	Test Acc
0	0.09	0.09	98.15	97.88	97.78
0.0001	0.09	0.09	98.15	97.88	97.78
0.01	0.10	0.10	98.15	97.88	97.78