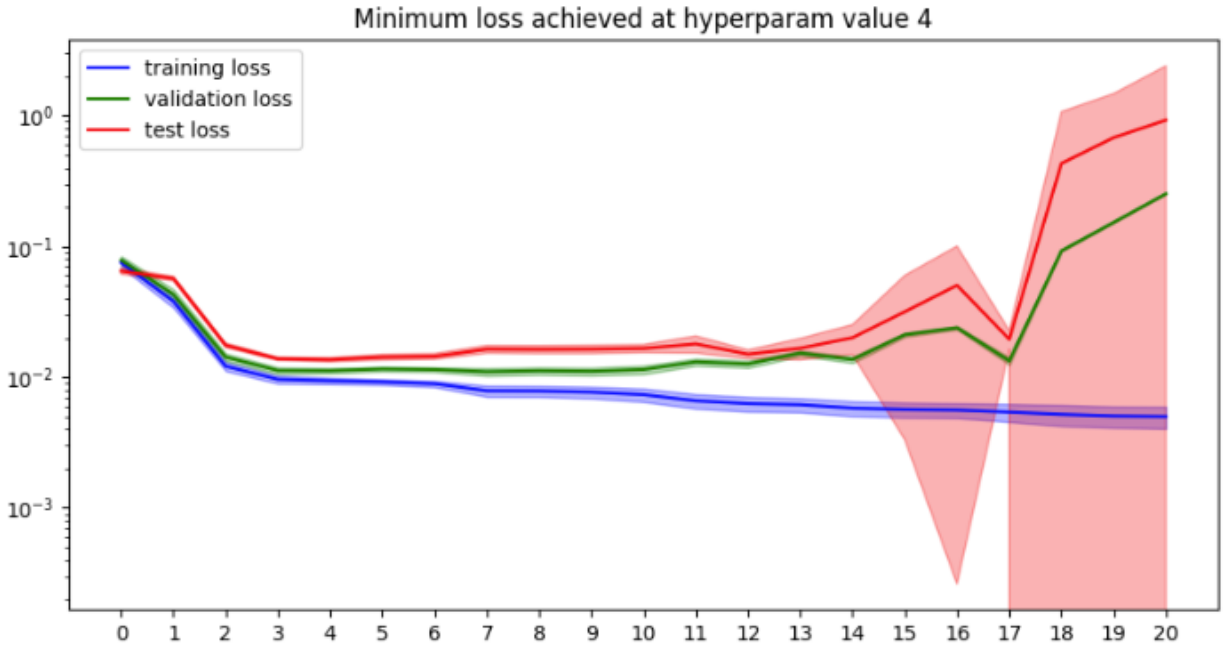
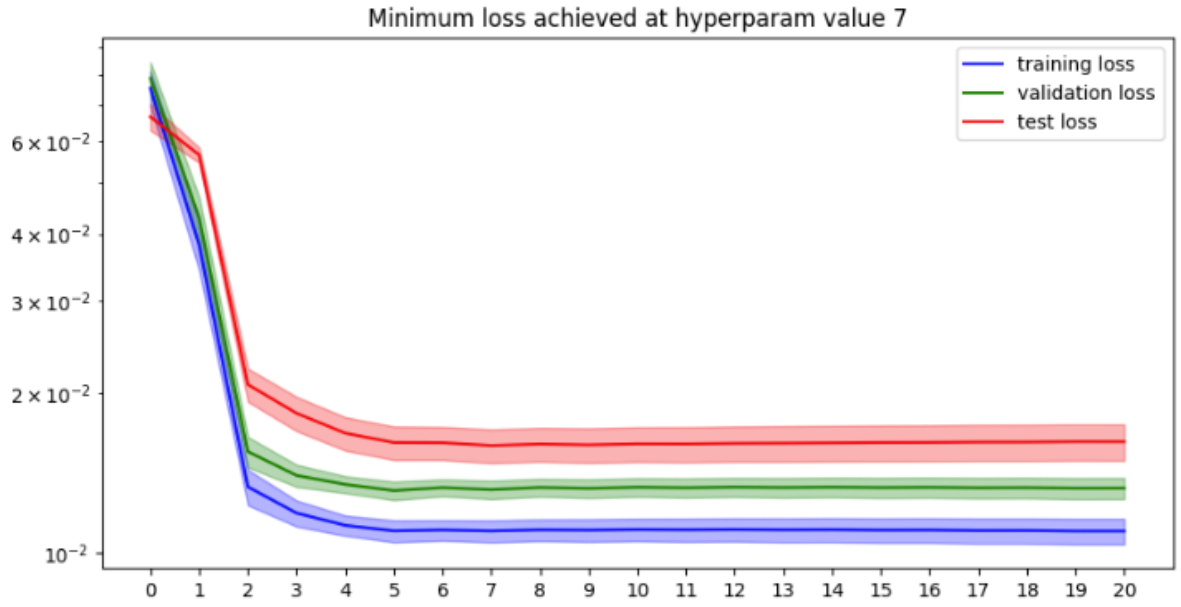
Q1



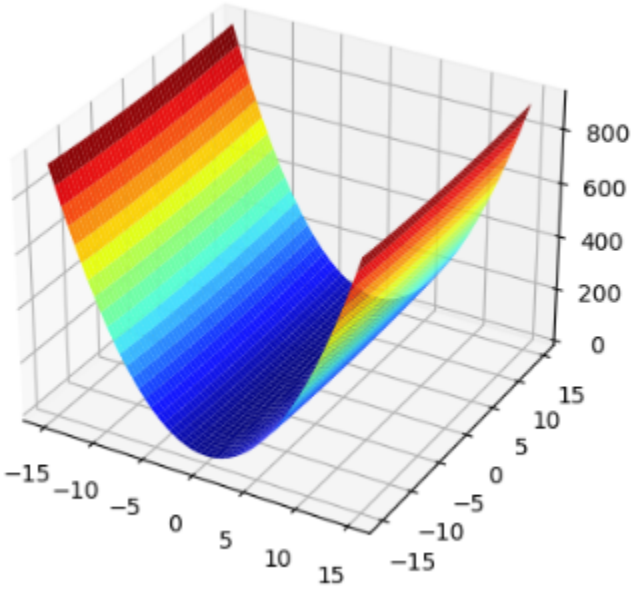
increase polynomial degree decreases training loss. Also reduce test loss and validation loss to min in degree 3, then increase beyond degree 3. Before degree 3, model underfitting data, after degree 3 overfitting data.



Regularization penalizes the model weights from having large values by adding the squared magni-tudes of weights to the loss function. Therefore, we achieve the best loss with a higher polynomial degree due to more generalization; in fact, coefficients of the higher order terms tend

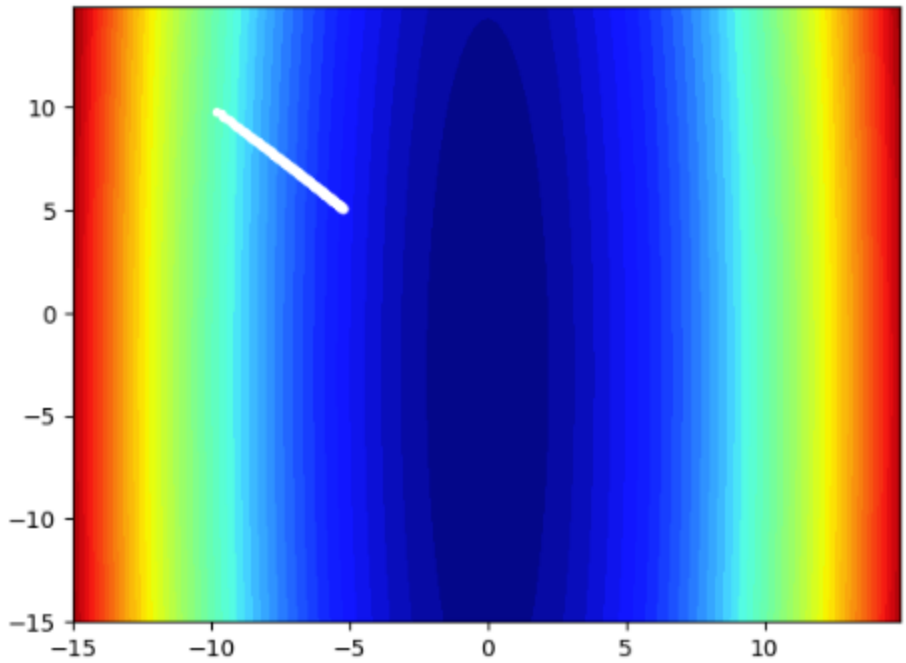
to be very small. The standard deviations of the three losses are also much smaller compared to the modelwithout regularization. However, the regularization method requires the choice of an additional hyperparameter to train the model

Q3

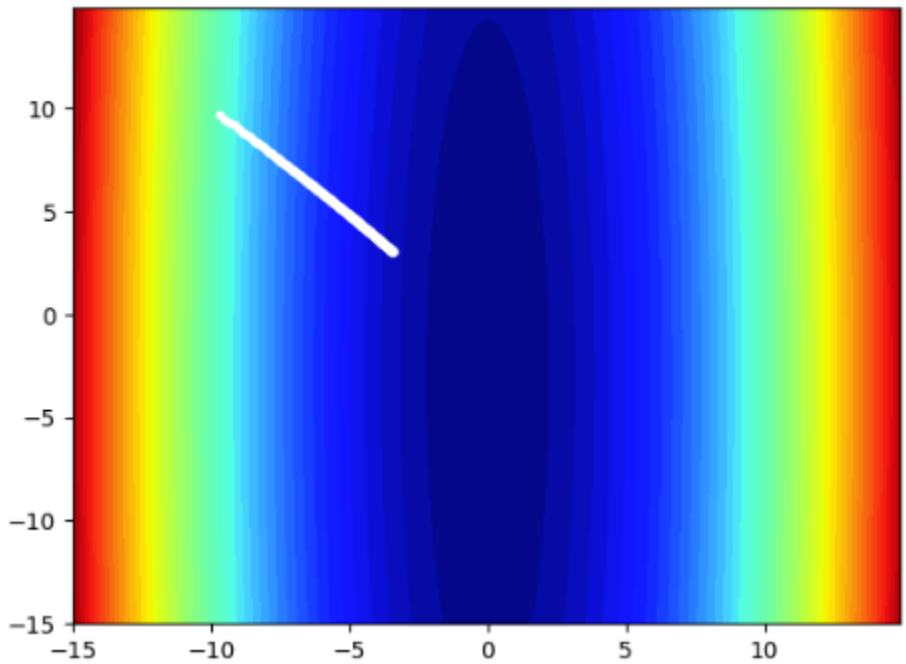


change step size to 0.3, is too large, gradient descent cant find optimal

Adagrad 0.2

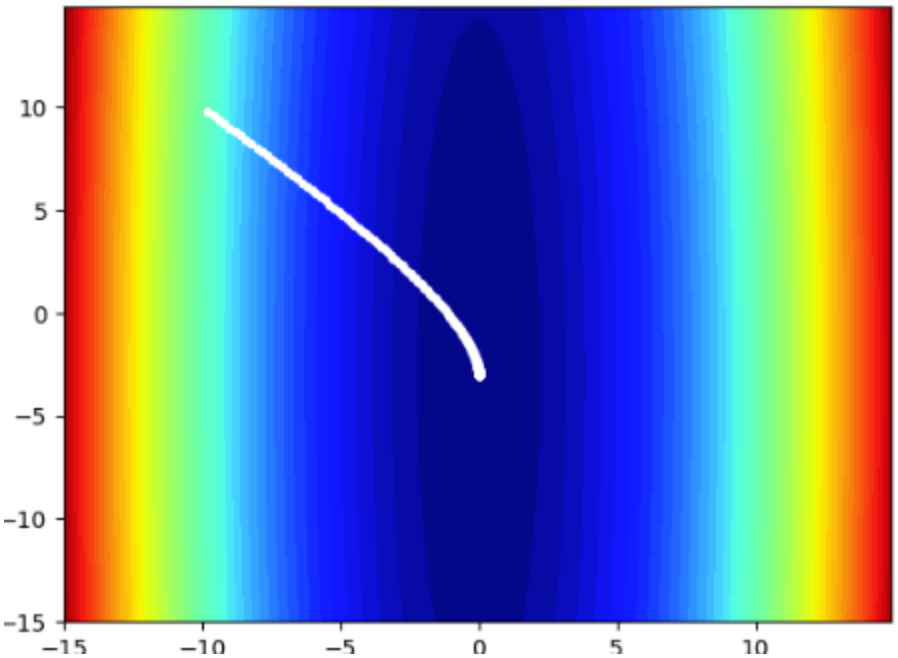


Adagrad 0.3



Using the step size of 0.3, one can observe that the convergence is still slow. Using agreater value for the step size will fix this problem

Adam



yes convert to min (-3, 0)