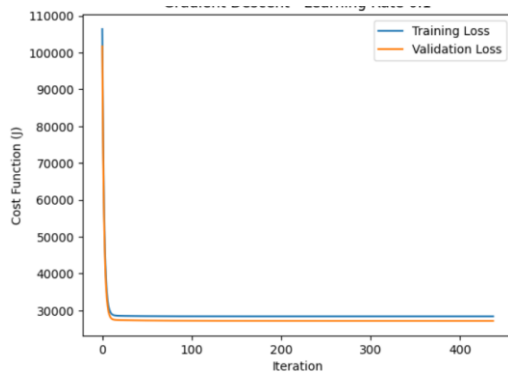


Note: I have considered each batch as 1 epoch, hence for example 16000 epochs with batch size 1 mean dataset is covered twice

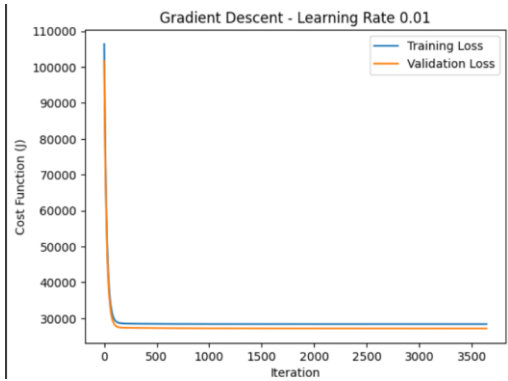
Normal Gradient Descent: Time ~ 0 second

Learning rate = 0.1, Epochs = 436



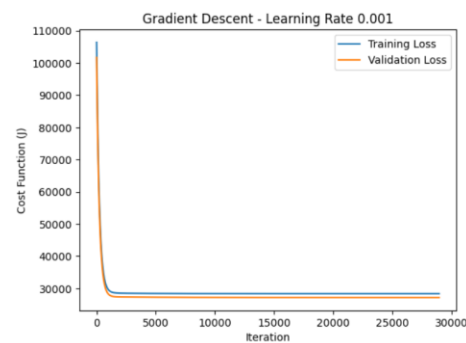
Slope : 241.5632, Constant: 132.336, Training Cost: 28386.341, CV Cost: 27158.154

Learning rate = 0.01, Epochs= 3639



Slope: 241.47146, Constant: 132.44347, Training Cost: 28386.344, CV Cost: 27158.464

Learning rate = 0.001, Epochs= 28945

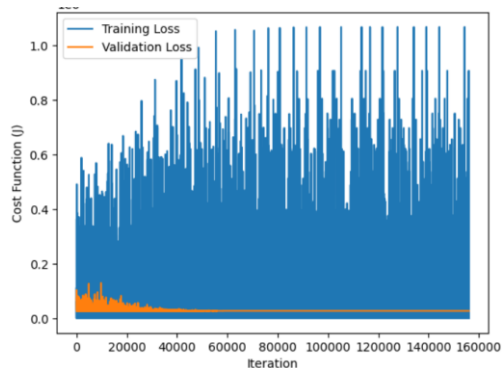


Slope: 241.1834, Constant: 132.7810, Training Cost: 28386.3732, CV Cost: 27159.457

Stochastic Gradient Descent

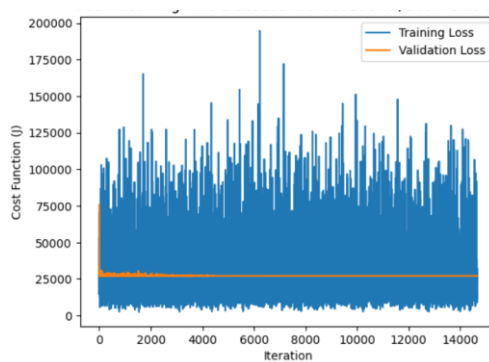
{Learning rate, batch size}

{0.1,1}



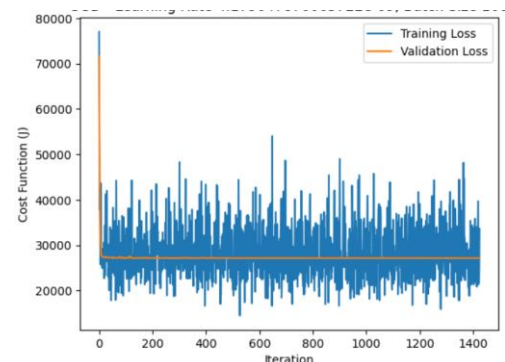
{0.01,1}

{0.1,10}

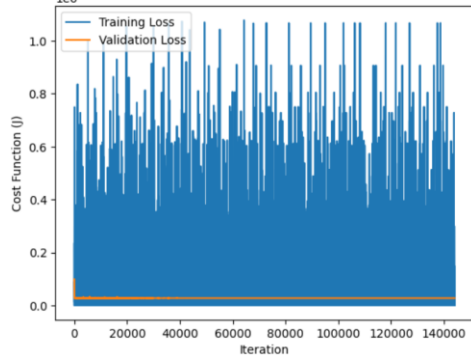


{0.01,10}

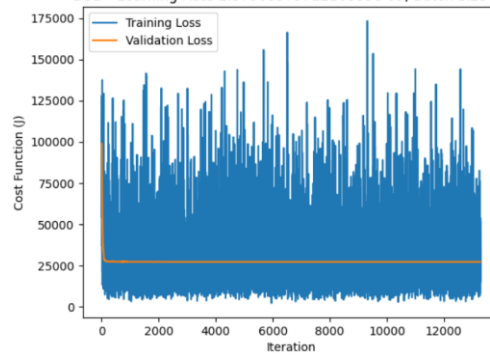
{0.1,100}



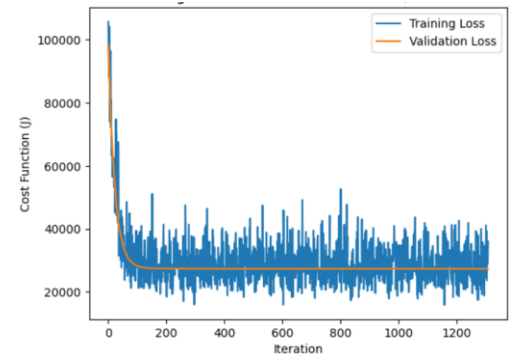
{0.01,100}



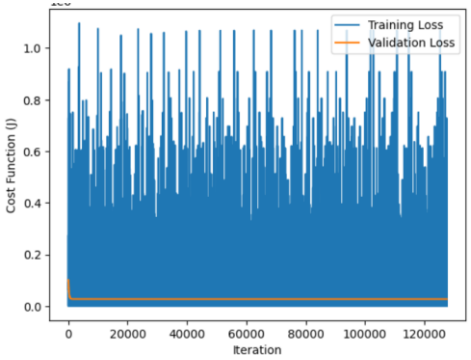
{0.001,1}



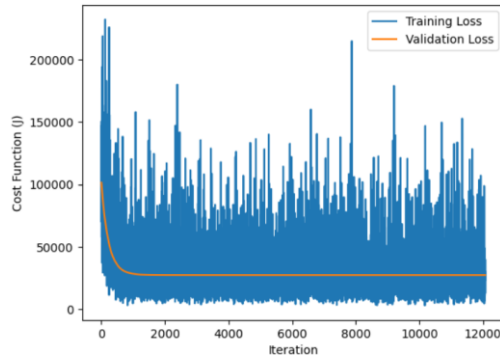
{0.001,10}



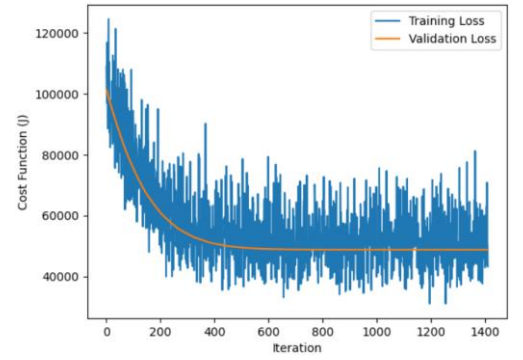
{0.001,100}



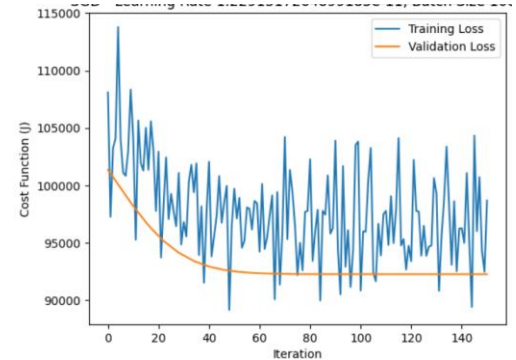
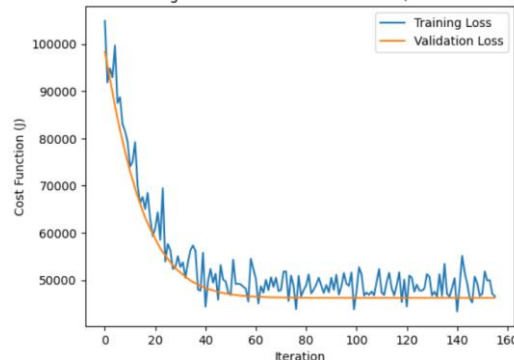
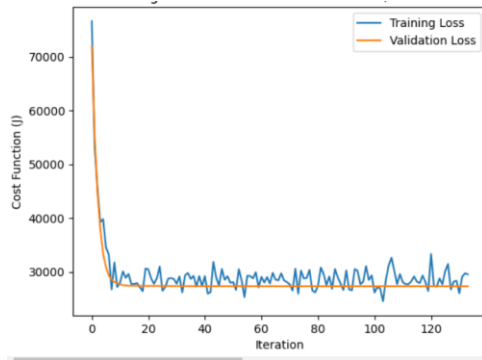
{0.1,1000}



{0.01,1000}



{0.001,1000}



Learning Rate	Batch Size	Time (seconds)	Epochs	Slope	Constant	Training loss	Validation loss
0.1	1	120	208024	241.857	131.614	2084.12	27155.15
0.1	10	13	19224	240.875	133.109	16728.92	27160.36
0.1	100	1	1864	241.646	132.399	33616.14	27158.78
0.1	1000	0	165	218.966	158.530	29537.69	27324.55
0.01	1	120	192024	242.123	132.162	15582.96	27159.10
0.01	10	13	17624	240.081	132.690	15190.52	27158.63
0.01	100	1	1704	217.781	159.180	36051.25	27331.32
0.01	1000	0	197	102.912	84.506	46575.23	46228.90
0.001	1	120	176024	241.294	132.504	2929.18	27158.24
0.001	10	11	16024	216.541	160.004	23729.55	27340.45
0.001	100	1	1784	96.169	78.960	43225.09	48768.78
0.001	1000	0	197	14.407	12.089	98692.86	92262.89

P.S.: all the timings are in Google collab, as shared in notebook. Also, if you think something is missing in run1/run2, you will find that in the notebook. K=24.

