

Learning Rate	No. of hidden layers	Size of each hidden layer	$\hat{\lambda}$ (regulariser)	RMSE(train)	RMSE(dev)
0.001	1	64	0	12.27	10.26
0.001	1	64	5	1427.54	1427.54
0.001	1	128	0	10.53	10.33
0.001	1	128	5	1427.54	1427.54
0.001	2	64	0	10.91	10.71
0.001	2	64	5	1427.54	1427.54
0.001	2	128	0	NaN	NaN
0.001	2	128	5	NaN	NaN
0.01	1	64	0	10.91	10.71
0.01	1	64	5	1427.54	1427.54
0.01	1	128	0	10.45	10.21
0.01	1	128	5	1427.54	1427.54
0.01	2	64	0	NaN	NaN
0.01	2	64	5	NaN	NaN
0.01	2	128	0	NaN	NaN
0.01	2	128	5	NaN	NaN