Jiafei Kuang

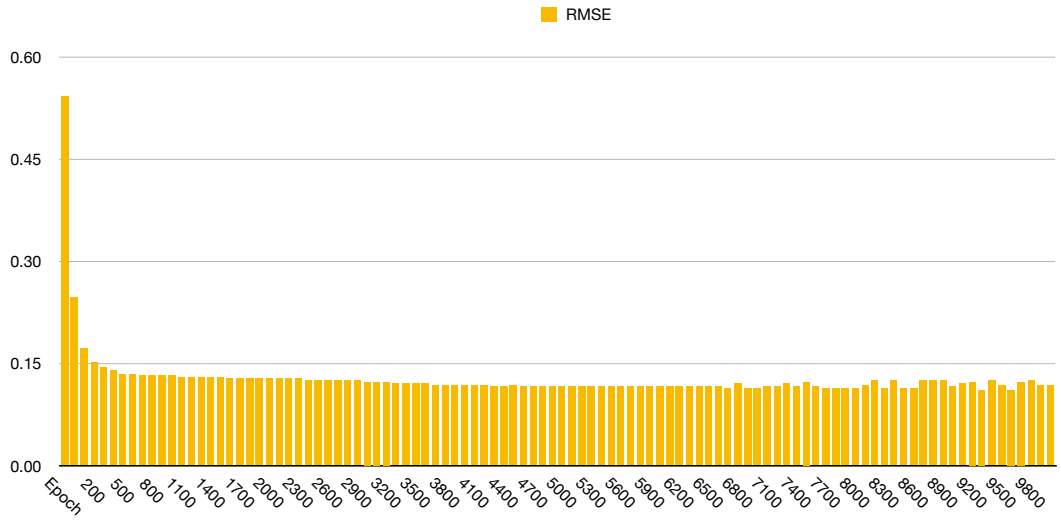
Assignment Five

1. run_iris()

This is the run_iris() with three hidden layers of four neurodes, 10001 epochs, verbosity 2, and 70% training factor.
Training final RMSE is 0.118855
Training RMSE is as below:

Iris

Epoch	RMSE	Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.5415625725815030	2600	0.12586117933248000	5100	0.1152892090834970	7600	0.1221173296099200
100	0.246597168051811	2700	0.12554358752458300	5200	0.11529191944919400	7700	0.11525383360292900
200	0.172221820665466	2800	0.12516748577548800	5300	0.11522279407433200	7800	0.11439022902522000
300	0.151506756439201	2900	0.124714490421484	5400	0.11517483650626600	7900	0.11429200181927600
400	0.143960146388326	3000	0.12416560499842200	5500	0.1151384851044780	8000	0.1141695023944860
500	0.138047215366367	3100	0.12350314922827300	5600	0.1151088087951180	8100	0.11377647633445300
600	0.134349766725291	3200	0.1227169860342970	5700	0.1150847686616500	8200	0.11706079721018600
700	0.133235523906122	3300	0.12182674072481000	5800	0.11506554021357900	8300	0.12569212367335700
800	0.13254643927746	3400	0.12092557215379100	5900	0.11505041704705300	8400	0.11336561805830900
900	0.131975527346094	3500	0.12016719423996900	6000	0.1150387862355690	8500	0.12512305744325300
1000	0.13148254410464400	3600	0.11961666499626800	6100	0.11503010374289700	8600	0.11220675330245100
1100	0.13103768517202700	3700	0.11923840625264200	6200	0.1150238759486190	8700	0.11409674762858300
1200	0.13060883308769300	3800	0.11900403142703600	6300	0.11501964503811400	8800	0.12569558697780100
1300	0.13015920094875500	3900	0.11882304724652800	6400	0.11501697690586900	8900	0.12387902778597600
1400	0.12966024168567600	4000	0.11849178681098900	6500	0.1150154507042240	9000	0.12433109737277300
1500	0.12912733543117000	4100	0.11776645989706900	6600	0.11501464940922600	9100	0.11580810631921100
1600	0.1286254156270340	4200	0.11717651148305400	6700	0.1150142065505620	9200	0.1205454423550090
1700	0.1282035597127140	4300	0.11716680127436600	6800	0.11375595252523100	9300	0.12331468098198500
1800	0.12785904213667800	4400	0.11589061246098000	6900	0.11930358421831600	9400	0.11114422793742100
1900	0.12756736722960900	4500	0.11630057742723200	7000	0.1144543889253710	9500	0.12400857749636500
2000	0.12730837719123600	4600	0.11691176759108300	7100	0.11431687104763300	9600	0.11825866873736800
2100	0.12706937216703500	4700	0.11474194893059700	7200	0.11449872833734900	9700	0.11120138812147600
2200	0.1268413045986100	4800	0.11565957787815800	7300	0.11482984651503000	9800	0.1221745350945120
2300	0.12661593059172900	4900	0.11591406823795600	7400	0.12040651773903600	9900	0.12442536932969700
2400	0.12638443488690600	5000	0.11523849424224200	7500	0.11468851904197700	10000	0.11855093986468800
2500	0.12613677745942200						



Testing RMSE is 0.116576823161683 77

This is the case of Iris run with three hidden layer of four neurodes, 10001 epochs, verbosity 2, and 70% training factor.

Initial (Epoch 0), the RMSE was relatively high at 0.54, indicating that the model's predictions were quite different from the actual values. However, as the training progresses, the RMSE consistently decreased at each epoch. After 100 epochs, the RMSE had dropped significantly to 0.25, and it continued to decrease further with each subsequent epoch.

After around 2000 epochs, the rate of decrease in RMSE slowed down, indicating that the model was starting to converge to a good solution. The network appeared to plateau around epoch 4300, with the RMSE fluctuating around 0.11. At around epoch 6800, the RMSE reached its minimum value of 0.113, indicating that the model is learning and improving over time.From epoch 4500-5200, 6900-7600, 8200-9500, the RMSE was even increasing.

There are several possible reasons for it:
1.Overfitting. One possible reason for this plateau could be overfitting. Overfitting occurs when a model becomes too complex and starts to fit the noise in the training data, rather than the underlying patterns. As a result, the model performs well on the training data, but not that good on

the test data. In this case, it's possible that the model is fitting the training data too closely and is unable to generalize to new data.

2. Model architecture. The complexity of the model architecture can also impact the performance during training. If the model is too simple, it may not be able to capture the underlying patterns in the data, while if the model is too complex, it may overfit to the training data. At this point, I will change number of the hidden level in the following training.

3. Learning rate. Moreover we may set a high learning rate and will lower the learning rating in the following training to see whether it matters.

4. Number of epochs. The optimal number of epochs may different depending on the models and the dataset. In this case, the model seems to have converged after around 2000 epochs, but it continues to improve slowly until epoch 6800. It is also possible that training the model for too long can lead to overfitting and worse performance on unseen data.

Overall, the network was able to reduce the RMSE significantly, which means that it could learn the underlying patterns in the data and also make accurate predictions. In the following training, I will modify the topology and learning rate of the networks, and the number of epochs used to train to see how these changes affect the results.

a. Change the number of hidden layers

This is the run_iris() with two hidden layers of four neurodes, 10001 epochs, verbosity 2, and 70% training factor.
Training Final Epoch RMSE = 0.27740470719987204
Training RMSE is as below(delete data from epoch 6001-10000):

Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.55474713967767	2100	0.2831001369593153	4100	0.28016734286163836
100	0.47125879177151	2200	0.28284855841118584	4200	0.2800846661276528
200	0.47111512560025	2300	0.282615691107184	4300	0.280005045018827
300	0.324598247442896	2400	0.2823994522007704	4400	0.27992825789351533
400	0.306458108680658	2500	0.2821980567936039	4500	0.27985410080994505
500	0.29906686712008	2600	0.2820099679458497	4600	0.2797823859610056
600	0.295026850046467	2700	0.281833855572119	4700	0.2797129402832051
700	0.29247526659671	2800	0.2816685625774492	4800	0.27964560421500295
800	0.290712001208714	2900	0.2815130769626834	4900	0.2795802305844535
900	0.289406069695317	3000	0.28136650886609194	5000	0.27951668360979065
1000	0.2883810673555466	3100	0.2812280716804703	5100	0.2794548379994158
1100	0.28753924197494607	3200	0.28109706652351485	5200	0.2793945781399772
1200	0.28682517813890174	3300	0.2809728694564629	5300	0.2793357973628899
1300	0.28620602795659056	3400	0.28085492094693465	5400	0.2792783972809893
1400	0.2856610701432564	3500	0.2807427171588319	5500	0.2792222871879991
1500	0.28517631471217664	3600	0.2806358027262854	5600	0.2791673835142878
1600	0.2847416902835747	3700	0.28053376473136377	5700	0.2791136093330313
1700	0.2843495464248261	3800	0.28043622765774684	5800	0.27906089391138134
1800	0.2839938329256498	3900	0.28034284913607055	5900	0.2790091723016775
1900	0.28366963090960406	4000	0.28025331633253797	6000	0.27895838496807535
2000	0.2833728687399323				

Testing RMSE is 0.31771681847237 63
In this iris run, the number of hidden layers is changed to two. The model is improving over time as the RMSE decreases with each epoch. The initial

RMSE was 0.55, which is relatively high, but after 100 epochs, it had already decreased to 0.47. After 1000 epochs, the RMSE had decreased to 0.29, and after 5000 epochs, it was down to 0.28. The final RMSE is 0.27740470719987204, which is higher than the model with three hidden layers. The RMSE does not increase during the whole epochs.

This is the run_iris() with five hidden layers of four neurodes, 10001 epochs, verbosity 2, and 70% training factor.
Training Final Epoch RMSE = 0.112368318520116
Training RMSE is as below(delete data from epoch 6000-10000):

LAYER 5

Epoch	RMSE	Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.744231842043	1500	0.115818816539	3000	0.112215489316	4500	0.111624482969
100	0.303306918490	1600	0.115449554683	3100	0.112131026094	4600	0.111614773652
200	0.162839729139	1700	0.115080399717	3200	0.112056840161	4700	0.111607022451
300	0.140891507425	1800	0.114722321977	3300	0.111991603300	4800	0.111601067293
400	0.131020307851	1900	0.114381726799	3400	0.111934192409	4900	0.111596763000
500	0.124259793240	2000	0.114062571824	3500	0.111883656621	5000	0.111593979268
600	0.120421045731	2100	0.113767765571	3600	0.111839188425	5100	0.111592598945
700	0.118825156436	2200	0.113499404258	3700	0.111800099214	5200	0.111592516562
800	0.117935546514	2300	0.113258471954	3800	0.111765798993	5300	0.111593637073
900	0.117478316300	2400	0.113044645233	3900	0.111735779711	5400	0.111595874758
1000	0.117218109654	2500	0.112856443291	4000	0.111709601656	5500	0.111599152280
1100	0.117003596525	2600	0.112691621290	4100	0.111686882351	5600	0.111603399848
1200	0.116769273153	2700	0.112547599630	4200	0.111667287498	5700	0.111608554484
1300	0.116492701074	2800	0.112421791883	4300	0.111650523576	5800	0.111614559379
1400	0.116172217364	2900	0.112311797996	4400	0.111636331769	5900	0.111621363318

Looking at the data, we can see that the RMSE decreased over time. At epoch 0, the RMSE was 0.75, which was relatively high, but by epoch 100, it had decreased to 0.30. This suggests that the model was making more accurate predictions as it learned from the data. After 1000 epochs, the RMSE had decreased to 0.117, The final RMSE was 0.112, which was lower than the model with three hidden layers, indicating the RMSE was improved as number of layers increase. The running time was a little bit slower than the model with three hidden layers.

Overall, changing the number of hidden layers is a trade-off between model complexity and performance. It may lead to faster training and reduced overfitting, but at the cost of decreased accuracy if the removed layer was providing important representational power.

b. Changing the Learning rate

In this run, the iris function was trained at the learning rate of 0.5, 0.7 and 0.9, with three hidden layers of four neurodes, 10001 epochs and verbosity 2.

The data of RMSE and the delta between each epoch are shown as below:

Epoch	RMSE of LR 0.5	RMSE of LR 0.7	RMSE of LR 0.9	Delta 0.5	Delta 0.7	Delta 0.9
0	0.67680992030	0.54156257258	0.64144868165			
100	0.32620975018	0.24659716805	0.33591642391	0.35060017012	0.21138675101	0.30553225774
200	0.31091073656	0.17222182067	0.33105792264	0.01529901363	0.01994386176	0.00485850127
300	0.30112922161	0.15150675644	0.31405701866	0.00978151494	0.00870804342	0.01700090398
400	0.29526762224	0.14396014639	0.30562713570	0.00586159938	0.00438760977	0.00842988296
500	0.29131654533	0.13804721537	0.30138344081	0.00395107691	0.00264158909	0.00424369489
600	0.28843449064	0.13434976673	0.29877762074	0.00288205468	0.00177903623	0.00260582007
700	0.28621528169	0.13323552391	0.29695460042	0.00221920896	0.00129278654	0.00182302032
800	0.28442851836	0.13254643928	0.29557309071	0.00178676333	0.00099447133	0.00138150971
900	0.28293843304	0.13197552735	0.29446706269	0.00149008531	0.00080084572	0.00110602802
1000	0.28166403681	0.13148254410	0.29354567023	0.00127439623	0.00066924703	0.00092139246
1100	0.28055506604	0.13103768517	0.29275586645	0.00110897078	0.00057582274	0.00078980377
1200	0.27957833360	0.13060883309	0.29206496420	0.00097673244	0.00050667016	0.00069090225
1300	0.27871043414	0.13015920095	0.29145163914	0.00086789945	0.00045350163	0.00061332506
1400	0.27793388291	0.12966024169	0.29090112392	0.00077655123	0.00041127264	0.00055051522
1500	0.27723502407	0.12912733543	0.29040261134	0.00069885885	0.00037682794	0.00049851258
1600	0.27660285091	0.12862541563	0.28994783448	0.00063217315	0.00034812451	0.00045477686
1700	0.27602830477	0.12820355971	0.28953026657	0.00057454614	0.00032378733	0.00041756791
1800	0.27550383096	0.12785904214	0.28914464425	0.00052447381	0.00030285521	0.00038562232
1900	0.27502307710	0.12756736723	0.28878666215	0.00048075386	0.00028463338	0.00035798209
2000	0.27458067351	0.12730837719	0.28845276249	0.00044240359	0.00026860570	0.00033389966
2100	0.27417206386	0.12706937217	0.28813998152	0.00040860965	0.00025438031	0.00031278097
2200	0.27379336917	0.12684130460	0.28784583307	0.00037869468	0.00024165456	0.00029414846
2300	0.27344127596	0.12661593059	0.28756821805	0.00035209321	0.00023019128	0.00027761501
2400	0.27311294302	0.12638443489	0.28730535327	0.00032833295	0.00021980230	0.00026286478
2500	0.27280592355	0.12613677746	0.28705571472	0.00030701947	0.00021033660	0.00024963855

Epoch	RMSE of LR 0.5	RMSE of LR 0.7	RMSE of LR 0.9	Delta 0.5	Delta 0.7	Delta 0.9
2600	0.27251810021	0.12586117933	0.28681799239	0.00028782334	0.00020167156	0.00023772233
2700	0.27224763109	0.12554358752	0.28659105393	0.00027046913	0.00019370659	0.00022693846
2800	0.27199290498	0.12516748578	0.28637391534	0.00025472610	0.00018635824	0.00021713859
2900	0.27175250464	0.12471449042	0.28616571723	0.00024040034	0.00017955659	0.00020819812
3000	0.27152517652	0.12416560500	0.28596570546	0.00022732813	0.00017324247	0.00020001177
3100	0.27130980604	0.12350314923	0.28577321535	0.00021537048	0.00016736535	0.00019249010
3200	0.27110539741	0.12271698603	0.28558765865	0.00020440863	0.00016188174	0.00018555670
3300	0.27091105699	0.12182674072	0.28540851271	0.00019434041	0.00015675388	0.00017914594
3400	0.27072597976	0.12092557215	0.28523531151	0.00018507723	0.00015194879	0.00017320120
3500	0.27054943802	0.12016719424	0.28506763815	0.00017654174	0.00014743748	0.00016767336
3600	0.27038077213	0.11961666500	0.28490511847	0.00016866589	0.00014319434	0.00016251968
3700	0.27021938266	0.11923840625	0.28474741569	0.00016138947	0.00013919664	0.00015770279
3800	0.27006472384	0.11900403143	0.28459422578	0.00015465883	0.00013542409	0.00015318991
3900	0.26991629790	0.11882304725	0.28444527357	0.00014842594	0.00013185856	0.00014895220
4000	0.26977365027	0.11849178681	0.28430030936	0.00014264763	0.00012848374	0.00014496421
4100	0.26963636536	0.11776645990	0.28415910595	0.00013728491	0.00012528497	0.00014120341
4200	0.26950406287	0.11717651148	0.28402145614	0.00013230249	0.00012224899	0.00013764981
4300	0.26937639452	0.11716680127	0.28388717052	0.00012766835	0.00011936380	0.00013428563
4400	0.26925304115	0.11589061246	0.28375607548	0.00012335337	0.00011661851	0.00013109504
4500	0.26913371011	0.11630057743	0.28362801159	0.00011933104	0.00011400322	0.00012806388
4600	0.26901813290	0.11691176759	0.28350283209	0.00011557721	0.00011150890	0.00012517950
4700	0.26890606305	0.11474194893	0.28338040156	0.00011206985	0.00010912730	0.00012243053
4800	0.26879727419	0.11565957788	0.28326059478	0.00010878886	0.00010685089	0.00011980678
4900	0.26869155830	0.11591406824	0.28314329574	0.00010571589	0.00010467277	0.00011729904
5000	0.26858872409	0.11523849424	0.28302839669	0.00010283421	0.00010258659	0.00011489904
5100	0.26848859557	0.11528920908	0.28291579742	0.00010012852	0.00010058653	0.00011259928
5200	0.26839101072	0.11529191945	0.28280540445	0.00009758485	0.00009866724	0.00011039296
5300	0.26829582024	0.11522279407	0.28269713051	0.00009519048	0.00009682377	0.00010827394
9900	0.26526523980	0.12442536933	0.27916001132	0.00004990566	0.00005227599	0.00005694813
10000	0.26521577437	0.11855093986	0.27910366615	0.00004946543	0.00005174263	0.00005634517

According to the data above, that the final RMSE values were 0.265, 0.11855, and 0.2791 at learning rates of 0.5, 0.7, and 0.9, respectively. And the RMSE decreased over time. Learning rates of 0.7 appeared to perform better than those of 0.5 and 0.9. Moreover, the RMSE did not

show a significant change with variations in learning rates, which indicate that the model was somewhat insensitive to changes in the learning rate. This may be due to the fact that the learning rate was high, resulting in a relatively small change in the margin.

c. Change the number of epochs used to train

I tried several time to 5000 and 20000 epochs each time, recording RMSE at each 100 epochs.
Below, the RMSE values for two different runs, one trained for 5000 epochs and the other for 20000 epochs, are presented(partially).

Epoch 5000, 20000

Epoch	RMSE	RMSE						
0	0.67701240986	0.58212582838	2500	0.12404132181	0.29277230170	5000	0.11244080920	0.28955278249
100	0.31479901051	0.33863773191	2600	0.12307742739	0.29258332995	5100		0.28946016200
200	0.18526770996	0.33393969475	2700	0.12200852821	0.29240316002	5200		0.28936906420
400	0.15295323077	0.31192120055	2900	0.11948272051	0.29206586150	5600		0.28901857409
500	0.14868214954	0.30713243140	3000	0.11809921483	0.29190732778	5700		0.28893413534
600	0.14582186962	0.30420098403	3100	0.11685472807	0.29175474911	5800		0.28885086758
700	0.14321641108	0.30215556812	3200	0.11590290780	0.29160762394	5900		0.28876872577
900	0.13819962659	0.29941965780	3400	0.11473206778	0.29132801310	6100		0.28860765545
1000	0.13740699222	0.29845587273	3500	0.11436227491	0.29119478508	6200		0.28852865177
1100	0.13667257055	0.29766466334	3600	0.11407240784	0.29106551279	6300		0.28845062333
1200	0.13564073050	0.29700185332	3700	0.11383709891	0.29093991567	6400		0.28837353879
1300	0.13453722636	0.29643605174	3800	0.11364053712	0.29081774107	6500		0.28829736903
1400	0.13349308225	0.29594470319	3900	0.11347251302	0.29069876077	6600		0.28822208695
1500	0.13251332263	0.29551149080	4000	0.11332619215	0.29058276791	6700		0.28814766734
1600	0.13157576618	0.29512450575	4100	0.11319684314	0.29046957454	6800		0.28807408674
1700	0.13067178623	0.29477494378	4200	0.11308108943	0.29035900938	6900		0.28800132334
1800	0.12979758759	0.29445618951	4300	0.11297645454	0.29025091595	7000		0.28792935677
1900	0.12895225698	0.29416318253	4400	0.11288107705	0.29014515097	7100		0.28785816811
2000	0.12813522832	0.29389198156	4500	0.11279352680	0.29004158299	13000		0.28466721600
2100	0.12734006622	0.29363946409	4600	0.11271268306	0.28994009111	13100		0.28462585057
2200	0.12655312270	0.29340311741	4700	0.11263765203	0.28984056400	18500		0.28275235424
2300	0.12575557865	0.29318089104	4800	0.11256770949	0.28974289894	18600		0.28272292998
2400	0.12492594237	0.29297109085	4900	0.11250226030	0.28964700104	20000		0.28232640282

From the data above, the model seems to have converged after around 2000 epochs, but it continues to improve slowly until epoch 7000. There is little progress after that. Overall, increasing the number of epochs can improve the performance of the model, up to a certain point. After that point, further increasing the number of epochs may not result in significant improvements in RMSE, or may even lead to overfitting.

2. run_sin()

This is the run_sin() with three hidden layers of one neurode, 10001 epochs, verbosity 2, and 10% training factor.
Training final RMSE is 0.0413, testing RMSE is 0.0333028162840915.
Training RMSE is as below:

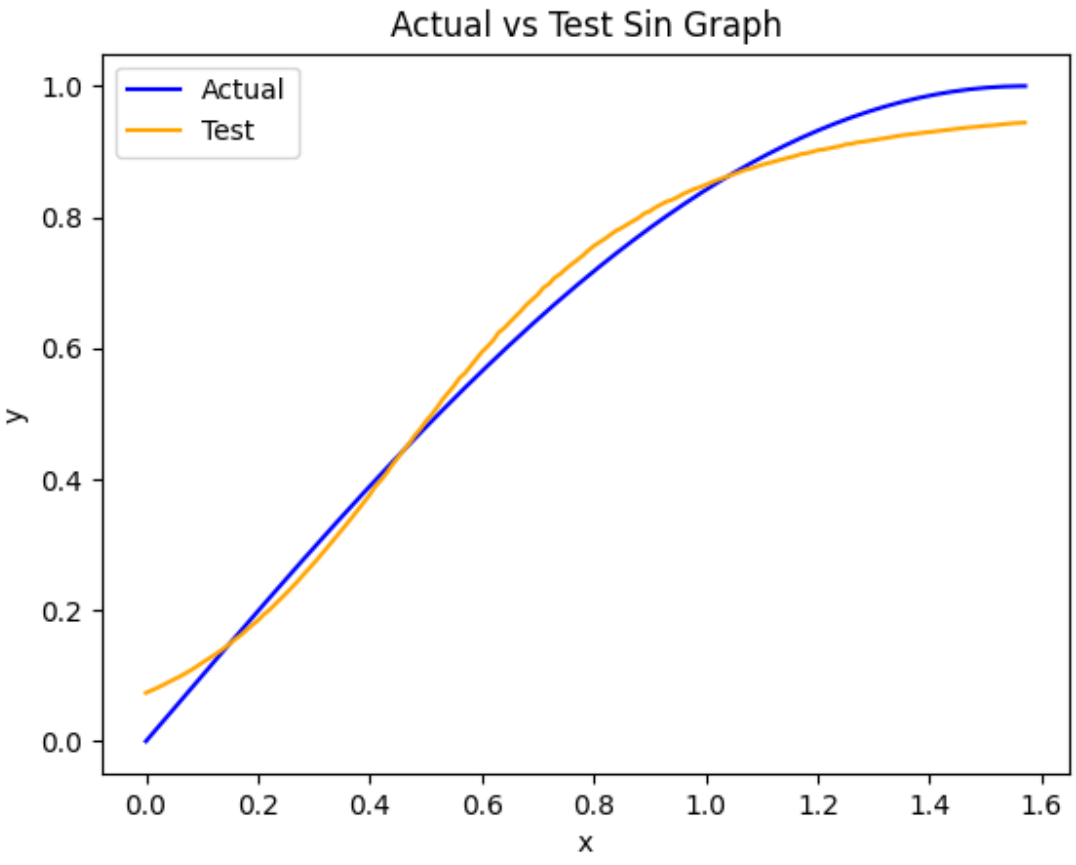
Epoch	RMSE	Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.33242964587	2300	0.30564309187	4400	0.08944488167	6500	0.05272990483
100	0.32243063186	2400	0.30538140056	4500	0.08551262214	6600	0.05205976104
200	0.31747010790	2500	0.30503308151	4600	0.08202453316	6700	0.05142872362
300	0.31458592179	2600	0.30453832156	4700	0.07891193559	6800	0.05083365008
400	0.31270180554	2700	0.30378209421	4800	0.07611973114	6900	0.05027170684
500	0.31137451579	2800	0.30252129015	4900	0.07360314574	7000	0.04974033358
600	0.31039138100	2900	0.30018272799	5000	0.07132537614	7100	0.04923721224
700	0.30963688310	3000	0.29526715106	5100	0.06925586033	7200	0.04876024018
800	0.30904218486	3100	0.28391927781	5200	0.06736898644	7300	0.04830750670
900	0.30856334914	3200	0.26057295740	5300	0.06564311484	7400	0.04787727252
1000	0.30817082499	3300	0.22873469654	5400	0.06405982729	7500	0.04746795180
1100	0.30784388959	3400	0.19895485060	5500	0.06260334343	7600	0.04707809631
1200	0.30756748719	3500	0.17502092631	5600	0.06126006186	7700	0.04670638147
1300	0.30733031947	3600	0.15631226166	5700	0.06001819554	7800	0.04635159398
1400	0.30712362892	3700	0.14154023361	5800	0.05886747921	9100	0.04287525955
1500	0.30694038157	3800	0.12965637787	5900	0.05779893250	9600	0.04192580854
1600	0.30677468280	3900	0.11991268512	6000	0.05680466651	9700	0.04175378024
1700	0.30662132346	4000	0.11178636387	6100	0.05587772480	9800	0.04158703249
1800	0.30647538348	4100	0.10490854953	6200	0.05501195148	9900	0.04142531283
1900	0.30633182924	4200	0.09901405415	6300	0.05420188139	10000	0.04126838438
2000	0.30618503216	4300	0.09390815915	6400	0.05344264785		

From the initial epochs, we can observe that the RMSE values are relatively high, indicating that the model's predictions have a high degree of error. As the model continues to train and adjust its parameters, the RMSE values start to decrease gradually.

Around epoch 1000, the RMSE values start to level off, indicating that the model's performance is stabilizing. From epoch 1000 to epoch 2000, the RMSE values continue to decrease but at a slower rate compared to the initial epochs.

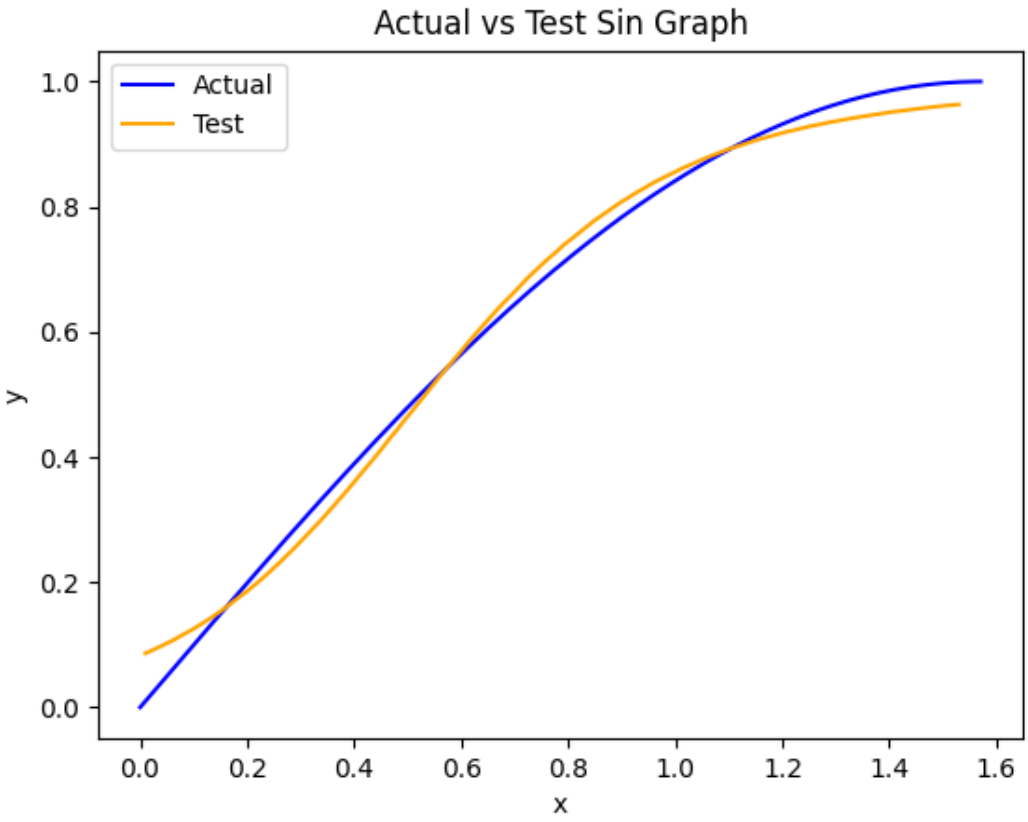
From epoch 2200 to epoch 3400, the RMSE values continue to decrease, but the rate of decrease is increased, indicating that the model's performance is improving significantly. From epoch 9000 to epoch 10000, the RMSE values appear to have plateaued, indicating that the model's performance has reached its optimal level.

Testing RMSE is 0.0333028162840915.



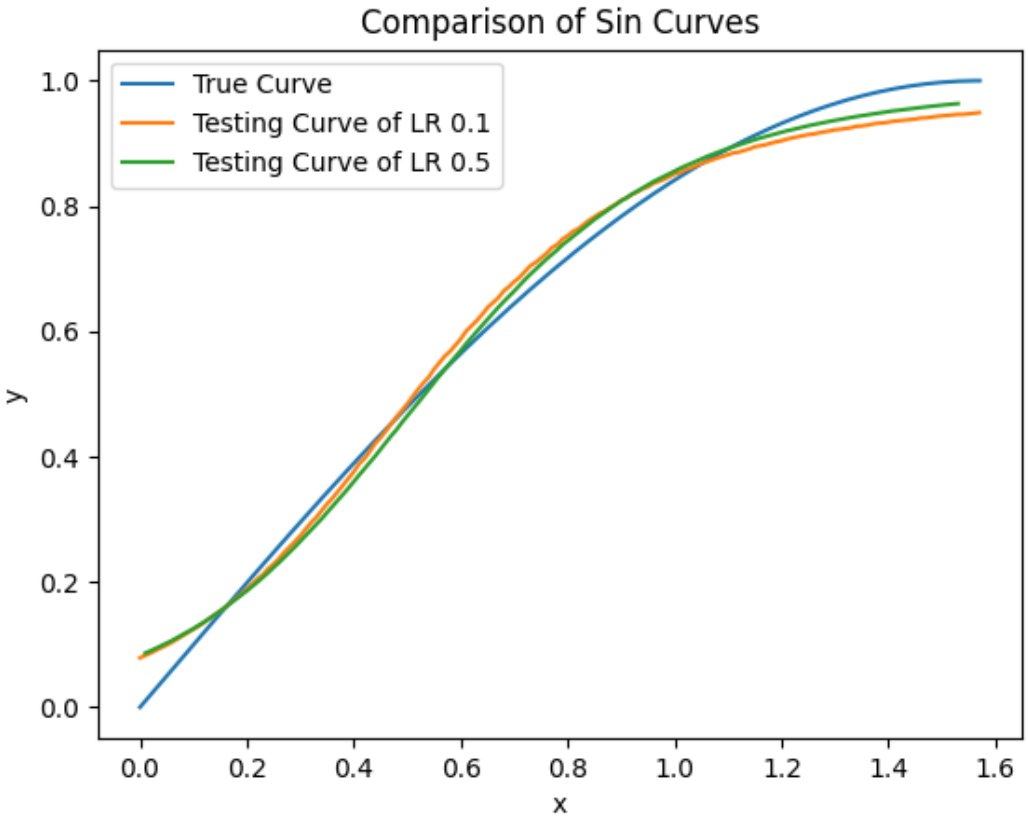
The image displayed above depicts the sin curve generated by the test method overlaid on an authentic sin wave graph spanning from 0 to pi/2. It appears that the sin curve being fitted to testing data within the range of x values between 0.158 and 1.064. The fitting is good within this range and matches the actual curve. However, it seems that the fitting is not as accurate or does not match the actual curve as well for other intervals outside of this range. The testing RMSE is reported as 0.0333028162840915, which may suggest that the overall accuracy of the sin curve fitting is relatively good.

As the learning rate used in the previous run was 0.1, I intend to increase it slightly to determine whether it could lead to an improved outcome.



In this run, the learning rate is 0.5. As the training final RMSE is 0.02682 and testing RMSE is 0.0271466586353512 27, both of them are lower than previous run. It indicate that as the learning rate increases, the RMSE of the regression model tends to decrease.

The image presented below displays a sine curve that has been generated by the test method using a learning rate of 0.5 and overlaid on top of an authentic sine graph ranging from 0 to pi/2.It seems that the sin curve has been fitted to the testing data, specifically within the x-value range of 0.148 to 1.148, resulting in improved performance compared to the previous run. The fitting of the curve within this range is accurate and closely resembles the actual curve.



The image depicted above exhibits a sin curve produced by the test method, with learning rates of 0.1 and 0.5, superimposed onto a sin graph ranging from 0 to pi/2. It is evident from the image that the learning rate of 0.5 provides a superior fit compared to 0.1, as the curve aligns more closely with the actual graph.

3. run_XOR()

This is the run_XOR() with three hidden layers of two neurodes, 10001 epochs, verbosity 2, and 100% training factor.

In the first run, I set the features and labels as below:

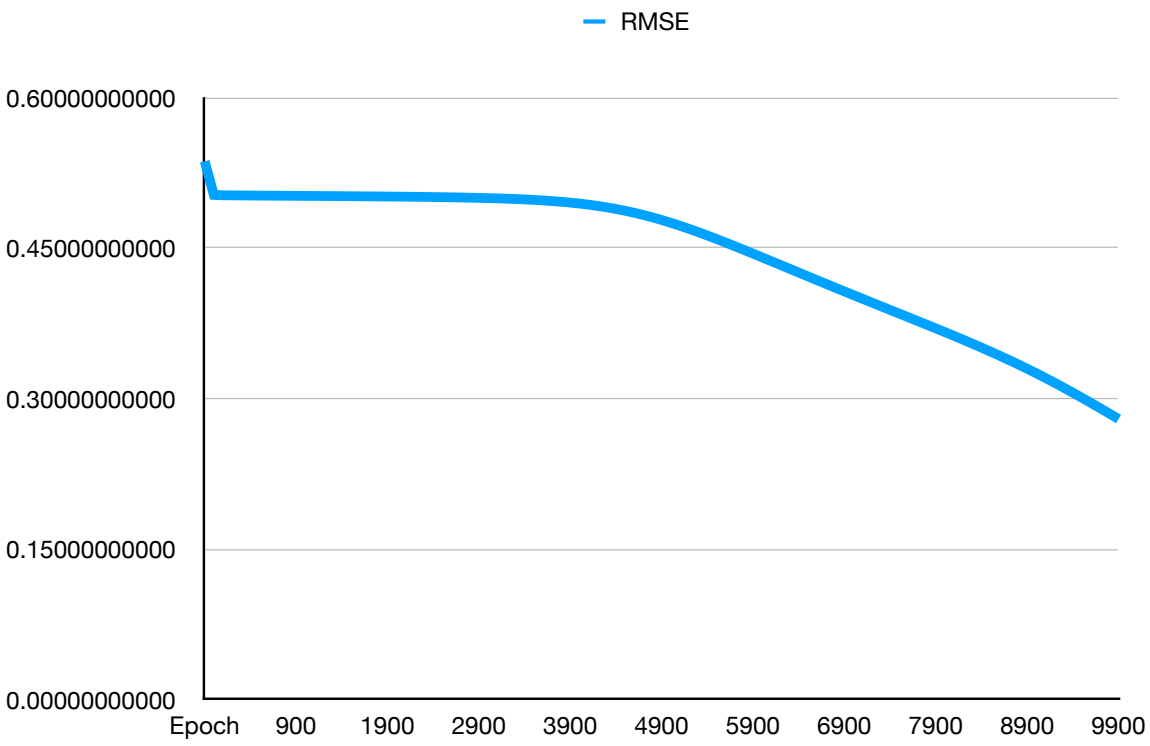
```
XOR_X = [[0, 0], [0, 1], [1, 0], [1, 1]]
XOR_Y = [[0], [1], [1], [0]]
```

The training final RMSE is 0.06535747654154249
Training RMSE is as below(partial):

Epoch	RMSE	Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.53637416319	4100	0.49428554993	7100	0.40306778549	9100	0.32522950916
100	0.50289639367	4200	0.49319083263	7200	0.39937179663	9200	0.32057918518
200	0.50251305923	4300	0.49193944685	7300	0.39568891305	9300	0.31581036537
300	0.50243519274	4400	0.49051715197	7400	0.39201663838	9400	0.31092571094
400	0.50236295607	4500	0.48891149755	7500	0.38835166219	9500	0.30593092566
500	0.50229344294	4600	0.48711267703	7600	0.38468986526	9600	0.30083478894
600	0.50222612412	4700	0.48511429135	7700	0.38102632425	9700	0.29564903396
700	0.50216056316	4800	0.48291391647	7800	0.37735531844	9800	0.29038807032
800	0.50209635200	4900	0.48051338846	7900	0.37367034224	9900	0.28506856562
900	0.50203309963	5000	0.47791876024	8000	0.36996412813	10000	0.27970891471
1000	0.50197042445						

Looking at the RMSE values in the table above, it is clear that the RMSE is decreasing as the number of epochs increases. However, it is also easy to see that the RMSE's decreasing speed is not constant.

Below is the picture of the training progress(epochs on the x-axis, RMSE on the y-axis). In the following graph, in the first 100 epochs, the RMSE decreases rapidly from 0.536 to 0.502. This is a sharp decrease, and it suggests that the model is learning quickly from the data.



The rate of decrease in RMSE gradually slows down between epochs 100 and 5000, with an average decrease speed of 0.0009. Although there is still a decrease in RMSE, it is less steep when compared to the initial 100 epochs. However, there appears to be an increase in the rate of decrease in RMSE after epoch 5000. From epoch 5000 to epoch 7500, the average decrease speed increases to 0.0036, and from epoch 7500 to epoch 10000, the average decrease speed further increases to 0.00435.

It probably because of several reasons. One possibility reason is that the model is learning more from the data as it is exposed to more epochs. As the rate of decrease in RMSE is still relatively high, it would be better to continue training the model for more epochs to determine if the performance can be further enhanced. I will discuss about it in the next section.

Extra Credit Opportunity #1 - Bias Nodes

In this section, I will train the XOR function with and without a bias input. For each scenario: train five times to 20,000 epochs each time, recording RMSE at each 100 epochs; average the results (i.e. average the five values at epoch 100, epoch 200, etc).

Firstly, I trained the XOR function without a bias input and with three hidden layers of two neurodes, 20001 epochs, and 100% training factor for five times.
In the first run, I set the features and labels as below:

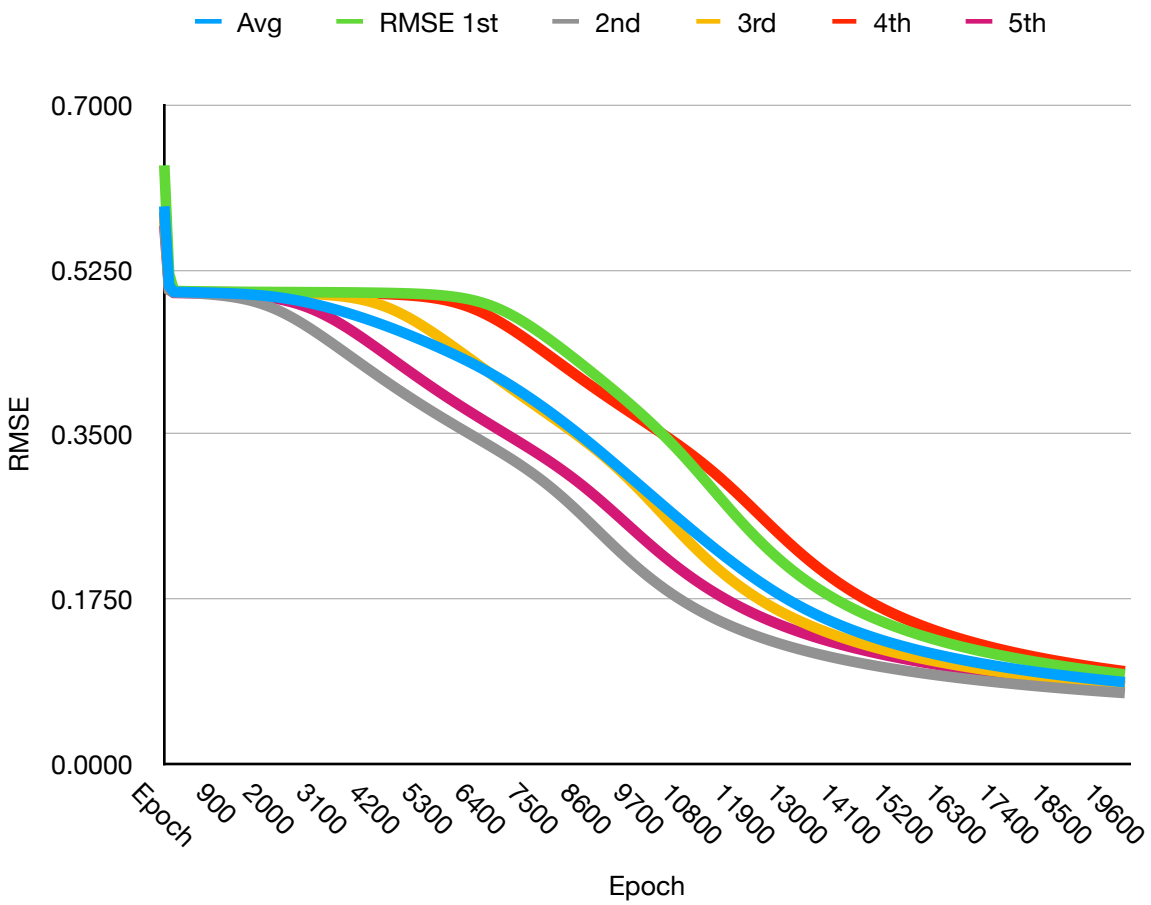
```
XOR_X = [[0, 0], [0, 1], [1, 0], [1, 1]]
XOR_Y = [[0], [1], [1], [0]]
```

Here is a partial dataset from five XOR sample runs, as well as the average results and a graph of each scenario's results plotted together.

Epoch	RMSE					
	1st	2nd	3rd	4th	5th	Avg
0	0.63629056710	0.58664698347	0.58138956363	0.58664427887	0.57233815878	0.59266191037
100	0.52117015112	0.50395070463	0.50430184853	0.50562824021	0.50378594433	0.50776737776
200	0.50284001538	0.50129616996	0.50202280971	0.50168872610	0.50064158534	0.50169786130
300	0.50258846736	0.50104317995	0.50193002837	0.50157722262	0.50053533623	0.50153484691
400	0.50247462415	0.50079967493	0.50186818654	0.50153737772	0.50047386526	0.50143074572
500	0.50237098435	0.50054103306	0.50181064245	0.50150231007	0.50040512631	0.50132601925
600	0.50227906516	0.50026185879	0.50175591986	0.50146943544	0.50032755252	0.50121876635
700	0.50219749473	0.49995644410	0.50170339250	0.50143829515	0.50023984192	0.50110709368
800	0.50212478749	0.49961841380	0.50165253422	0.50140862914	0.50014040774	0.50098895448
900	0.50205966258	0.49924058664	0.50160285293	0.50138022034	0.50002737404	0.50086213931
1000	0.50200103549	0.49881485039	0.50155387564	0.50135287374	0.49989853174	0.50072423340
3000	0.50137122019	0.46407875843	0.49968851399	0.50079514899	0.48493417519	0.49017356336
3100	0.50134471949	0.46076047303	0.49944522469	0.50075391038	0.48285426140	0.48903171780
3200	0.50131718042	0.45735745639	0.49916554353	0.50070959672	0.48059581567	0.48782911854
3300	0.50128836993	0.45388442479	0.49884314462	0.50066178556	0.47816332996	0.48656821097
3400	0.50125803253	0.45035503516	0.49847068246	0.50060999504	0.47556385820	0.48525152068
3500	0.50122588525	0.44678178108	0.49803969304	0.50055367331	0.47280658461	0.48388152346
3600	0.50119161180	0.44317597223	0.49754052134	0.50049218588	0.46990235077	0.48246052840
3700	0.50115485589	0.43954776967	0.49696229490	0.50042480066	0.46686319812	0.48099058385
3800	0.50111521331	0.43590625404	0.49629296769	0.50035067015	0.46370196376	0.47947341379

Epoch	RMSE					
	1st	2nd	3rd	4th	5th	Avg
6100	0.49602598332	0.35841780324	0.44047086619	0.48976536731	0.38264318082	0.43346464018
6200	0.49520727632	0.35540318865	0.43679636536	0.48821142518	0.37937463609	0.43099857832
6300	0.49426171249	0.35240015376	0.43310119666	0.48647943544	0.37613804364	0.42847610840
6400	0.49317407310	0.34940347417	0.42939236464	0.48456381946	0.37293113051	0.42589297237
6500	0.49192939569	0.34640733325	0.42567623593	0.48246217748	0.36975123112	0.42324527469
6600	0.49051371468	0.34340528830	0.42195853792	0.48017546944	0.36659528145	0.42052965836
6700	0.48891488084	0.34039024175	0.41824433640	0.47770796706	0.36345980770	0.41774344675
6800	0.48712337544	0.33735442177	0.41453799906	0.47506699109	0.36034091089	0.41488473965
6900	0.48513301522	0.33428937771	0.41084315163	0.47226247659	0.35723424904	0.41195245403
7000	0.48294144493	0.33118599683	0.40716263304	0.46930642747	0.35413501864	0.40894630418
7100	0.48055033815	0.32803455016	0.40349845463	0.46621232636	0.35103793787	0.40586672144
7200	0.47796527118	0.32482477585	0.39985176674	0.46299455874	0.34793723407	0.40271472132
7300	0.47519528923	0.32154600882	0.39622283468	0.45966789564	0.34482663866	0.39949173341
7400	0.47225223272	0.31818736530	0.39261102476	0.45624706141	0.34169939339	0.39619941551
7500	0.46914992352	0.31473798892	0.38901479950	0.45274639681	0.33854827203	0.39283947616
7600	0.46590331912	0.31118736202	0.38543172085	0.44917961422	0.33536562253	0.38941352775
7700	0.46252772894	0.30752568074	0.38185845945	0.44555963361	0.33214343459	0.38592298747
13400	0.19340917072	0.11926645429	0.14787757202	0.21720239903	0.13705914797	0.16296294881
13500	0.18988564600	0.11798208743	0.14573000615	0.21301896139	0.13541390599	0.16040612139
13600	0.18649149408	0.11673412928	0.14365702360	0.20895284594	0.13381656063	0.15793041071
13700	0.18322195974	0.11552106574	0.14165517958	0.20500559690	0.13226521428	0.15553380325
13800	0.18007223529	0.11434145956	0.13972119057	0.20117775120	0.13075805041	0.15321413741
13900	0.17703751675	0.11319394616	0.13785193104	0.19746898057	0.12929333121	0.15096914115
14000	0.17411304889	0.11207722961	0.13604442914	0.19387822602	0.12786939486	0.14879646571
14100	0.17129416053	0.11099007889	0.13429586167	0.19040382227	0.12648465285	0.14669371524
14200	0.16857629162	0.10993132424	0.13260354857	0.18704361052	0.12513758715	0.14465847242
14300	0.16595501360	0.10889985381	0.13096494704	0.18379503908	0.12382674744	0.14268832019
14400	0.16342604392	0.10789461041	0.12937764552	0.18065525183	0.12255074826	0.14078085999
14500	0.16098525597	0.10691458848	0.12783935757	0.17762116499	0.12130826626	0.13893372665
14600	0.15862868521	0.10595883122	0.12634791577	0.17468953290	0.12009803745	0.13714460051
14700	0.15635253221	0.10502642788	0.12490126577	0.17185700380	0.11891885455	0.13541121684
14800	0.15415316330	0.10411651124	0.12349746041	0.16912016663	0.11776956441	0.13373137320
14900	0.15202710925	0.10322825516	0.12213465405	0.16647558972	0.11664906548	0.13210293473
15000	0.14997106256	0.10236087235	0.12081109707	0.16391985258	0.11555630538	0.13052383799

19500	0.09807777244	0.07693874027	0.08505223777	0.10179574386	0.08430358333	0.08923361553
19600	0.09742356508	0.07656515362	0.08456729821	0.10105695028	0.08385722188	0.08869403781
19700	0.09678127736	0.07619686430	0.08409027226	0.10033289173	0.08341758989	0.08816377911
19800	0.09615056755	0.07583375136	0.08362095411	0.09962311524	0.08298452684	0.08764258302
19900	0.09553110694	0.07547569758	0.08315914516	0.09892718597	0.08255787735	0.08713020260
20000	0.09492257923	0.07512258934	0.08270465365	0.09824468632	0.08213749093	0.08662639989



From the data above, the RMSE has a rapid decrease at the first 100 epoch among each of five times. However, after that, the speed of decrease slows down considerably after 100 epoch. Although the duration of the sharp decrease varies among these five time periods, the underlying pattern remains consistent. It appears that there is an acceleration in the decrease of RMSE after a certain epoch. Specifically, in the first run, the rapid decrease occurred between epochs 6000 and 15000, in the second

run between epochs 2500 and 13000, in the third run between epochs 4500 and 13500, in the fourth run between epochs 6200 and 15000, and in the fifth run between epochs 4200 and 12000. After this period, the rate of decrease slows down, but the overall trend of decreasing remains the same across all five time periods. The average initial RMSE is 0.5927 and final RMSE is 0.08866, indicating the model has learned to make accurate predictions after being trained.

In addition, I trained the XOR function with a bias input and with three hidden layers of three neurodes, 20001 epochs, and 100% training factor for five times.
In the first run, I set the features and labels as below:

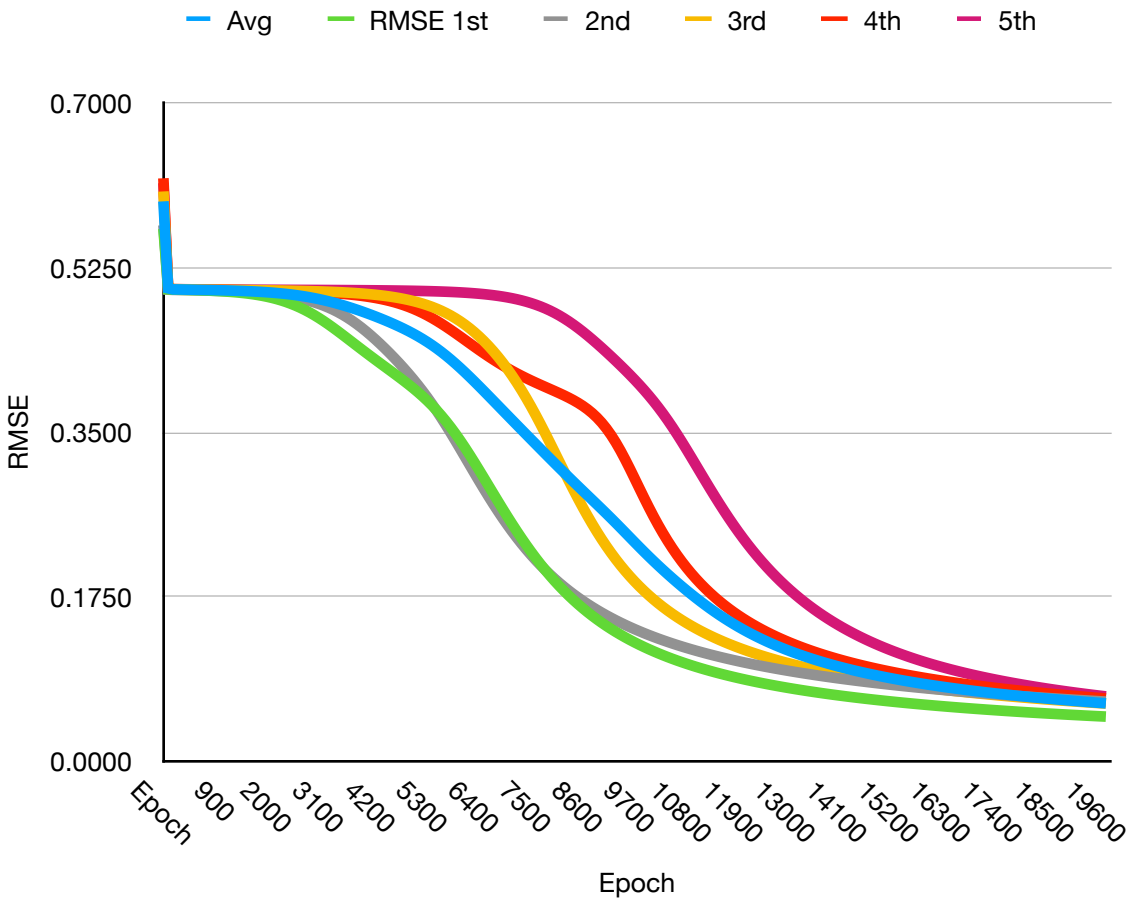
```
XOR_X = [[0, 0, 1], [0, 1, 1], [1, 0, 1], [1, 1, 1]]
XOR_Y = [[0], [1], [1], [0]]
```

Here is a partial dataset from five XOR sample runs, as well as the average results and a graph of each scenario's results plotted together.

Epoch	RMSE					
	1st	2nd	3rd	4th	5th	Avg
0	0.56653913237	0.57037390056	0.60625841981	0.62024957345	0.61516012970	0.59571623118
100	0.50217581267	0.50180210155	0.50422388784	0.50359515072	0.50433488711	0.50322636798
200	0.50182515143	0.50157969732	0.50236901175	0.50217560426	0.50236208268	0.50206230949
2000	0.49612856134	0.49830453951	0.50119249418	0.50096550299	0.50171208164	0.49966063593
2100	0.49524235977	0.49789167597	0.50112853683	0.50087487891	0.50169199753	0.49936588980
2200	0.49423423080	0.49743065381	0.50106151525	0.50077819150	0.50167241099	0.49903540047
2300	0.49309120346	0.49691516544	0.50099090738	0.50067466016	0.50165320234	0.49866502776
2400	0.49179998167	0.49633806473	0.50091615277	0.50056342214	0.50163425728	0.49825037572
2500	0.49034716426	0.49569126658	0.50083664487	0.50044352017	0.50161546570	0.49778681232
2600	0.48871954953	0.49496564293	0.50075172264	0.50031388860	0.50159672045	0.49726950483
2700	0.48690456308	0.49415092042	0.50066066124	0.50017333752	0.50157791624	0.49669347970
2800	0.48489084845	0.49323558815	0.50056266168	0.50002053476	0.50155894851	0.49605371631
2900	0.48266904496	0.49220682872	0.50045683924	0.49985398547	0.50153971234	0.49534528214
3000	0.48023274282	0.49105049113	0.50034221038	0.49967200883	0.50152010130	0.49456351089
3100	0.47757955012	0.48975113097	0.50021767814	0.49947271169	0.50150000632	0.49370421545
3200	0.47471213947	0.48829215005	0.50008201556	0.49925395861	0.50147931443	0.49276391563
3300	0.47163908155	0.48665607299	0.49993384713	0.49901333816	0.50145790750	0.49174004946

Epoch	RMSE					
	1st	2nd	3rd	4th	5th	Avg
4700	0.41984746051	0.43929891660	0.49519607500	0.49094179936	0.50098951282	0.46925475286
4800	0.41617104958	0.43430948911	0.49450880650	0.48975033876	0.50093377287	0.46713469136
4900	0.41247967369	0.42915328936	0.49374241502	0.48842293874	0.50087290761	0.46493424488
5000	0.40874916875	0.42382412006	0.49288801226	0.48694711193	0.50080627566	0.46264293773
5100	0.40495122454	0.41830943829	0.49193590119	0.48531064870	0.50073314849	0.46024807224
5200	0.40105356092	0.41259140737	0.49087552315	0.48350215811	0.50065269766	0.45773506944
5300	0.39702022842	0.40664858380	0.48969540003	0.48151172817	0.50056398042	0.45508798417
5400	0.39281217635	0.40045815359	0.48838307463	0.47933168486	0.50046592341	0.45229020257
5500	0.38838825259	0.39399858886	0.48692505641	0.47695740690	0.50035730426	0.44932532181
5600	0.38370678508	0.38725252520	0.48530678465	0.47438812841	0.50023673100	0.44617819087
5700	0.37872781925	0.38020957921	0.48351262654	0.47162763917	0.50010261916	0.44283605667
5800	0.37341593188	0.37286876587	0.48152593185	0.46868477978	0.49995316633	0.43928971514
5900	0.36774332935	0.36524017755	0.47932916752	0.46557363340	0.49978632423	0.43553452641
6000	0.36169274200	0.35734566925	0.47690415258	0.46231334185	0.49959976823	0.43157113478
6100	0.35525955509	0.34921844174	0.47423240383	0.45892752048	0.49939086427	0.42740575708
6200	0.34845274770	0.34090158527	0.47129558493	0.45544330493	0.49915663331	0.42304997123
6300	0.34129450615	0.33244579500	0.46807602573	0.45189012087	0.49889371331	0.41852003221
6400	0.33381869836	0.32390656796	0.46455724836	0.44829830918	0.49859831883	0.41383582854
6500	0.32606859330	0.31534122833	0.46072440966	0.44469775381	0.49826619816	0.40901963665
6600	0.31809424140	0.30680610604	0.45656455603	0.44111664621	0.49789258799	0.40409482753
6700	0.30994984920	0.29835412500	0.45206659995	0.43758048365	0.49747216526	0.39908464461
6800	0.30169136382	0.29003296122	0.44722097373	0.43411135213	0.49699899579	0.39401112934
6900	0.29337437845	0.28188383021	0.44201899455	0.43072749871	0.49646647935	0.38889423626
7000	0.28505239842	0.27394087669	0.43645207095	0.42744316244	0.49586729072	0.38375115985
14500	0.06920455071	0.08726254403	0.09108456289	0.10643676692	0.14009204158	0.09881609323
14600	0.06858308466	0.08661665754	0.09018594819	0.10511897853	0.13739719417	0.09758037262
14700	0.06797535689	0.08598315274	0.08931019388	0.10384233947	0.13480197175	0.09638260295
14800	0.06738090804	0.08536165753	0.08845640542	0.10260491127	0.13230169126	0.09522111470
14900	0.06679929877	0.08475181471	0.08762373431	0.10140487361	0.12989191266	0.09409432681
15000	0.06623010876	0.08415328127	0.08681137520	0.10024051569	0.12756842658	0.09300074150
15100	0.06567293562	0.08356572768	0.08601856323	0.09911022832	0.12532724248	0.09193893947
15200	0.06512739405	0.08298883726	0.08524457155	0.09801249666	0.12316457713	0.09090757533
15300	0.06459311491	0.08242230553	0.08448870901	0.09694589360	0.12107684361	0.08990537333
15400	0.06406974442	0.08186583965	0.08375031799	0.09590907357	0.11906064078	0.08893112328
15500	0.06355694335	0.08131915787	0.08302877246	0.09490076698	0.11711274330	0.08798367679

19700	0.04852949122	0.06459895558	0.06262592998	0.06821960489	0.07105409815	0.06300561596
19800	0.04827781486	0.06430633990	0.06229463103	0.06781297074	0.07044194790	0.06262674089
19900	0.04802953206	0.06401722699	0.06196807549	0.06741291061	0.06984240742	0.06225403051
20000	0.04778457132	0.06373154958	0.06164615525	0.06701925626	0.06925509103	0.06188732469



Based on the provided data, it can be observed that the shape of the RMSE curve for both the models - with and without bias - is similar. The RMSE decreases rapidly during the first 100 epochs in all the five runs. However, after that, the speed of decrease slows down. Although the duration of the initial sharp decrease varies across the five runs, the overall trend remains consistent. There seems to be a significant acceleration in the RMSE decrease after a certain epoch.

In particular, the first run saw a rapid decrease between epochs 1000 and 11000, the second run between epochs 4000 and 11000, the third run between epochs 6000 and 11500, the fourth run between epochs 6000 and 12000, and the fifth run between epochs 7500 and 14000. After this period, the rate of decrease slows down, but the overall trend of decrease remains the same across all five runs.

The model's initial average RMSE is 0.5957, which indicates that the predictions of the model were not accurate in the beginning. However, after being trained, the final RMSE has reduced to 0.06189, implying that the model has learned to make more accurate predictions.

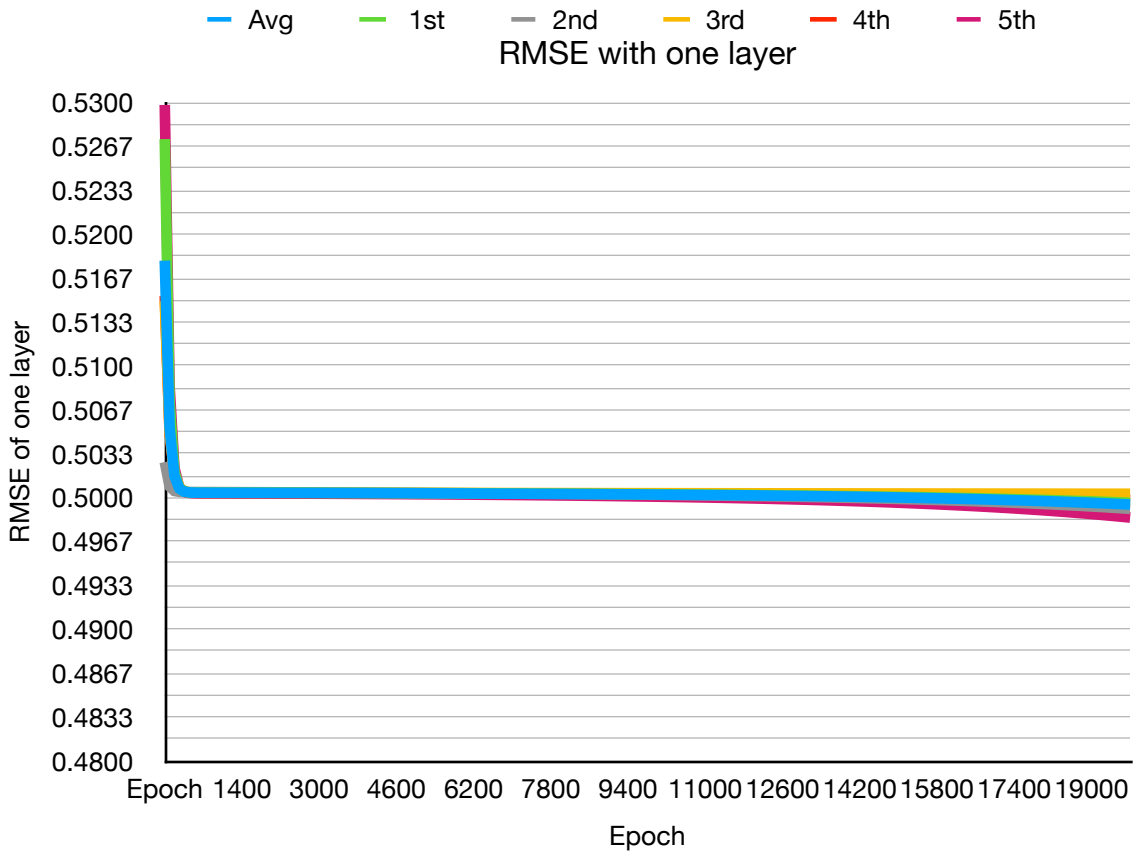
According the data in these two training process, the final RMSE without bias is 0.08866 and with bias is 0.06189. We can see that the model with bias has a lower RMSE value. This suggests that the model with bias is better at making predictions for the XOR problem than the model without bias. It may because the bias provides additional information to the network to adjust its predictions, enabling it to learn the XOR function more accurately.

Extra Credit Opportunity #2 - Vanishing Gradient

Deltas tend to get small as the error backpropagates. Networks with more layers train more slowly, in terms of epochs, than networks with fewer layers. Sometimes networks with many layers get “stuck."
In this section, I trained the XOR function(with bias) with one, three and five hidden layers. For each scenario: train five times to 20,000 epochs each time, recording RMSE at each 100 epochs; calculate average the results.

Firstly, I trained the XOR function(with bias) with one hidden layers of three neurodes, 20001 epochs, and 100% training factor for five times.
Here is a partial dataset from five XOR sample runs, as well as the average results and a graph of each scenario's results plotted together.

Epoch	RMSE with one layer					
	1st	2nd	3rd	4th	5th	Avg
0	0.52722066814	0.50268835078	0.51491903359	0.51530294294	0.52981706089	0.51798961127
100	0.50735653406	0.50083535279	0.50493880305	0.50494142568	0.50843265749	0.50530095461
200	0.50196198853	0.50049293060	0.50176969025	0.50168497484	0.50225102232	0.50163212131
300	0.50072228172	0.50042958465	0.50078770084	0.50072416047	0.50075841418	0.50068442837
400	0.50045218479	0.50041693251	0.50048785245	0.50044831166	0.50041939680	0.50044493564
500	0.50039441934	0.50041356760	0.50039793838	0.50036999331	0.50034393137	0.50038397000
600	0.50038197670	0.50041195190	0.50037171918	0.50034773152	0.50032689823	0.50036805550
700	0.50037899250	0.50041068160	0.50036440244	0.50034124619	0.50032251571	0.50036356769
800	0.50037790207	0.50040948123	0.50036246786	0.50033917770	0.50032076395	0.50036195856
10000	0.50025086909	0.50020933464	0.50032434730	0.50027607708	0.50007113399	0.50022635242
10100	0.50024835257	0.50020543323	0.50032390588	0.50027477221	0.50006570505	0.50022363379
10200	0.50024579636	0.50020147026	0.50032346413	0.50027344266	0.50006017794	0.50022087027
10300	0.50024319973	0.50019744450	0.50032302206	0.50027208784	0.50005455071	0.50021806097
10400	0.50024056194	0.50019335473	0.50032257967	0.50027070717	0.50004882138	0.50021520498
10500	0.50023788224	0.50018919967	0.50032213698	0.50026930003	0.50004298794	0.50021230137
10600	0.50023515985	0.50018497803	0.50032169399	0.50026786581	0.50003704831	0.50020934920
19600	0.49964750962	0.49923122591	0.50028135248	0.49988548647	0.49861187822	0.49953149054
19700	0.49963406970	0.49920857098	0.50028090723	0.49987582096	0.49857810986	0.49951549574
19800	0.49962036843	0.49918545576	0.50028046221	0.49986594885	0.49854370004	0.49949918706
19900	0.49960640031	0.49916187118	0.50028001743	0.49985586598	0.49850863935	0.49948255885
20000	0.49959215970	0.49913780805	0.50027957290	0.49984556811	0.49847291847	0.49946560545



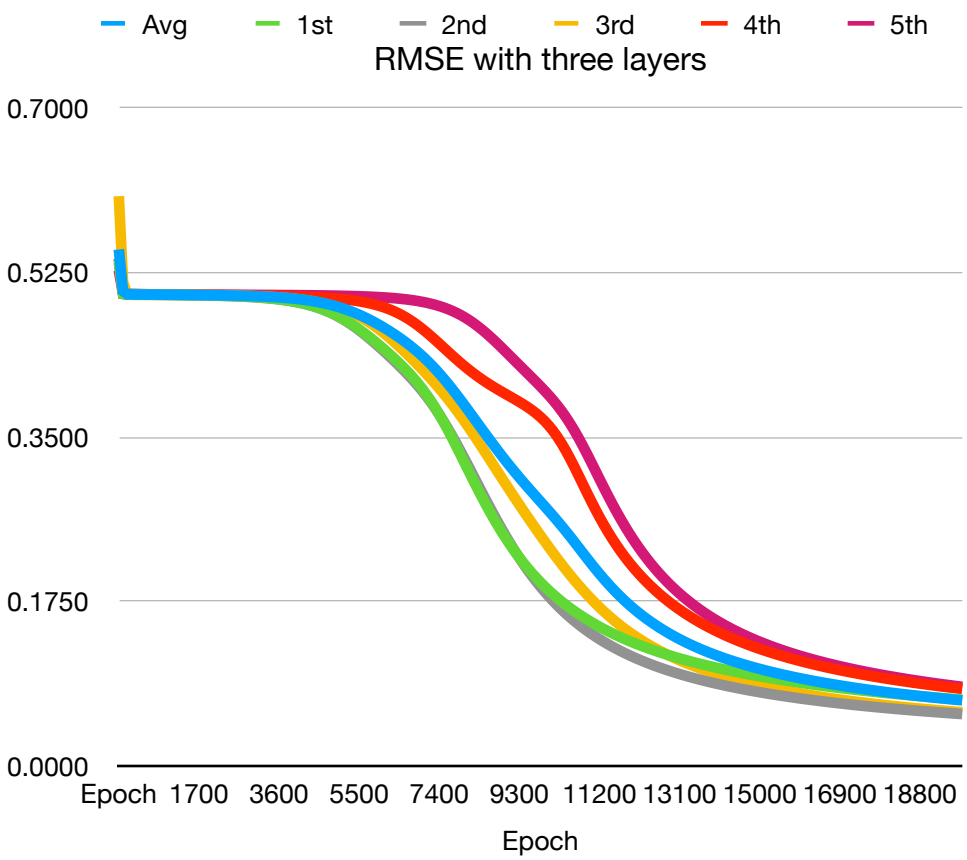
The consistency in the shape and values of the RMSE trend across five runs, as shown in both the table and graph, suggests that the model has successfully learned the underlying patterns in the training data, leading to consistent and same accurate predictions. The RMSE has a rapid decrease at the first 100 epoch, subsequently decreases at a much slower rate. Despite the decrease in the average initial RMSE from 0.51798 to an average final RMSE of 0.49946, the improvement in the model's performance may not be considered significant. Furthermore, the fact that the RMSE trends across the five runs are very similar suggests that the model's accuracy is not satisfactory when trained with only one layer. In the following, I will delve into the topic of using additional layers in neural networks.

In this scenario, I trained the XOR function(with bias) with three hidden layers of three neurodes, 20001 epochs, and 100% training factor for five times.

Here is a partial dataset from five XOR sample runs, as well as the average results and a graph of each scenario's results plotted together.

Epoch	RMSE with three layers					
	1st	2nd	3rd	4th	5th	Avg
0	0.53861294271	0.53500486582	0.60586360424	0.53956875744	0.52704367483	0.54921876901
100	0.50196765209	0.50335897736	0.51462579642	0.50314963285	0.50185193374	0.50499079849
200	0.50109193248	0.50220689396	0.50140485444	0.50181076052	0.50115010321	0.50153290892
300	0.50105640024	0.50202410377	0.50101895075	0.50170529725	0.50113809190	0.50138856878
400	0.50103215471	0.50188242380	0.50101028034	0.50164624481	0.50113504360	0.50134122945
5000	0.48259414401	0.48232417040	0.49017053615	0.49763628283	0.50017496001	0.49058001868
5100	0.48053029989	0.48020887626	0.48890877313	0.49729074649	0.50009727748	0.48940719465
5200	0.47828233486	0.47789194211	0.48750711616	0.49690329256	0.50001198759	0.48811933466
5300	0.47584726180	0.47536847230	0.48595240367	0.49646802279	0.49991814482	0.48671086108
5400	0.47322543961	0.47263748222	0.48423125563	0.49597820792	0.49981466815	0.48517741071
5500	0.47042086604	0.46970239312	0.48233052688	0.49542619071	0.49970032077	0.48351605950
5600	0.46744115018	0.46657117518	0.48023790009	0.49480328873	0.49957368677	0.48172544019
5700	0.46429709617	0.46325604280	0.47794260253	0.49409970237	0.49943314467	0.47980571771
5800	0.46100188290	0.45977266672	0.47543620215	0.49330443669	0.49927683731	0.47775840516
5900	0.45756988697	0.45613894478	0.47271340642	0.49240524861	0.49910263786	0.47558602493
6000	0.45401524984	0.45237344668	0.46977275891	0.49138863540	0.49890811150	0.47329164047
6100	0.45035032100	0.44849369813	0.46661711246	0.49023988473	0.49869047270	0.47087829780
6200	0.44658411049	0.44451448345	0.46325376411	0.48894321165	0.49844653760	0.46834842146
6300	0.44272086025	0.44044632494	0.45969416892	0.48748201198	0.49817267156	0.46570320753
6400	0.43875880836	0.43629425184	0.45595320496	0.48583926481	0.49786473172	0.46294205234
6500	0.43468918840	0.43205692141	0.45204803018	0.48399811555	0.49751800466	0.46006205204
7000	0.41186895948	0.40906761389	0.43063117664	0.47135448699	0.49498890248	0.44358222789
7100	0.40653672984	0.40392024271	0.42604285190	0.46810166252	0.49426964969	0.43977422733
7200	0.40083751229	0.39850903788	0.42135564622	0.46462437106	0.49345664760	0.43575664301
7300	0.39472837579	0.39279949184	0.41656064739	0.46094547930	0.49253804886	0.43151440863
7400	0.38817823350	0.38676226207	0.41164490587	0.45709569090	0.49150085592	0.42703638965
7500	0.38117302375	0.38037615967	0.40659272032	0.45311232907	0.49033093711	0.42231703398
7600	0.37371905214	0.37363062489	0.40138724272	0.44903740336	0.48901310438	0.41735748550
7700	0.36584376053	0.36652731924	0.39601228354	0.44491518257	0.48753127991	0.41216596516

Epoch	RMSE with three layers					
	1st	2nd	3rd	4th	5th	Avg
10000	0.19528932799	0.19052496333	0.24503432422	0.37095120104	0.40356073650	0.28107211062
10100	0.19079615808	0.18519474519	0.23873492387	0.36661991606	0.39885960453	0.27604106954
10200	0.18651687753	0.18011103612	0.23252430687	0.36171753782	0.39396672262	0.27096729619
10300	0.18243939287	0.17526465214	0.22641253009	0.35616478224	0.38881165388	0.26581860224
10400	0.17855209612	0.17064585838	0.22041080591	0.34990581260	0.38332053118	0.26056702084
10500	0.17484391785	0.16624459793	0.21453090638	0.34292272723	0.37742273921	0.25519297772
10600	0.17130435639	0.16205068112	0.20878452749	0.33524693885	0.37105953213	0.24968920720
10700	0.16792348892	0.15805393871	0.20318269069	0.32696300949	0.36419330647	0.24406328686
14500	0.10127042787	0.08426732404	0.09351396465	0.13654060617	0.14651558537	0.11242158162
14600	0.10035331796	0.08334987949	0.09227300523	0.13462716429	0.14420964856	0.11096260311
14700	0.09945806479	0.08245734656	0.09107153024	0.13277890971	0.14199031870	0.10955123400
14800	0.09858386430	0.08158871319	0.08990772736	0.13099258638	0.13985293522	0.10818516529
14900	0.09772995142	0.08074302016	0.08877988641	0.12926514316	0.13779314240	0.10686222871
15000	0.09689559774	0.07991935779	0.08768639299	0.12759371888	0.13580686766	0.10558038701
15100	0.09608010937	0.07911686290	0.08662572249	0.12597562841	0.13389030131	0.10433772490
15200	0.09528282496	0.07833471597	0.08559643446	0.12440835003	0.13203987765	0.10313244061
15300	0.09450311383	0.07757213851	0.08459716739	0.12288951372	0.13025225746	0.10196283818
15400	0.09374037423	0.07682839062	0.08362663377	0.12141689041	0.12852431182	0.10082732017
15500	0.09299403169	0.07610276867	0.08268361549	0.11998838211	0.12685310696	0.09972438098
15600	0.09226353757	0.07539460316	0.08176695953	0.11860201278	0.12523589043	0.09865260069
15700	0.09154836757	0.07470325678	0.08087557397	0.11725591994	0.12367007816	0.09761063928
15800	0.09084802044	0.07402812249	0.08000842422	0.11594834691	0.12215324260	0.09659723133
15900	0.09016201674	0.07336862182	0.07916452954	0.11467763567	0.12068310167	0.09561118109
16000	0.08948989765	0.07272420323	0.07834295972	0.11344222027	0.11925750866	0.09465135790
19300	0.07295537989	0.05740792648	0.05967246486	0.08576499231	0.08825135556	0.07281042382
19400	0.07257888356	0.05707103235	0.05927954315	0.08518968764	0.08762489980	0.07234880930
19500	0.07220774737	0.05673944221	0.05889351649	0.08462472374	0.08701037426	0.07189516081
19600	0.07184184980	0.05641302516	0.05851419650	0.08406980952	0.08640742324	0.07144926084
19700	0.07148107305	0.05609165460	0.05814140165	0.08352466481	0.08581570517	0.07101089986
19800	0.07112530291	0.05577520808	0.05777495693	0.08298901992	0.08523489197	0.07057987596
19900	0.07077442862	0.05546356714	0.05741469358	0.08246261508	0.08466466835	0.07015599455
20000	0.07042834275	0.05515661713	0.05706044883	0.08194520006	0.08410473123	0.06973906800



Based on the provided table and graph, it can be observed that RMSE decreases rapidly during the first 100 epochs in all the five runs. However, after that, the speed of decrease slows down. Although the duration of the initial sharp decrease varies across the five runs, the overall trend remains consistent. There seems to be a significant acceleration in the RMSE decrease after a certain epoch.

In particular, the first run saw a rapid decrease between epochs 5000 and 11500, the second run between epochs 5000 and 12000, the third run between epochs 6000 and 12000, the fourth run between epochs 7500 and 14000, and the fifth run between epochs 8500 and 15000. After this period, the rate of decrease slows down, but the overall trend of decrease remains the same across all five runs.

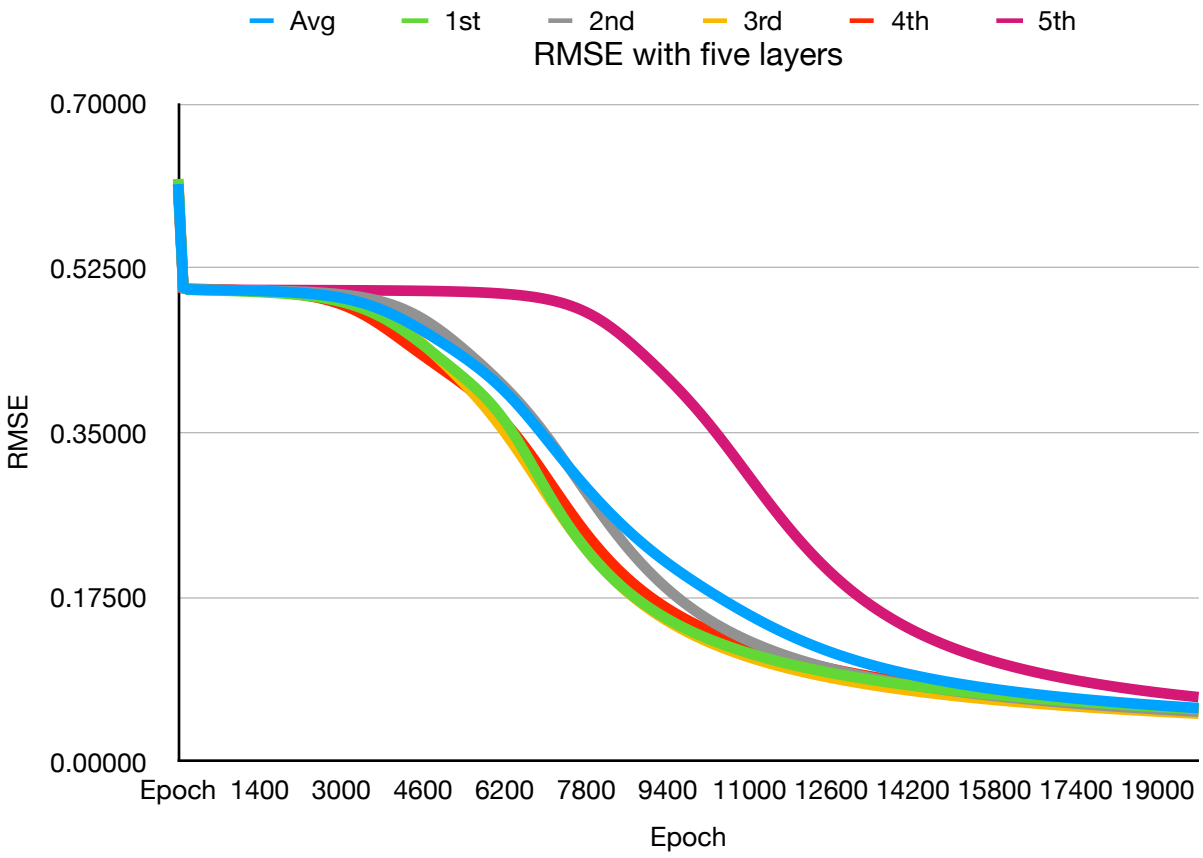
The average initial RMSE is 0.5492, and the average final RMSE is 0.0697. When compared to the model with only one layer, the five runs of the model with multiple layers exhibit a greater diversity in the shape of their respective lines. In the case of the model with one single layer, the RMSE

decreased from 0.51798 to 0.49946. However, for the model with three layers, the RMSE decreased from 0.5492 to 0.0697, indicating that the reduction in error is more pronounced when utilizing a greater number of layers.

Finally, I trained the XOR function(with bias) with five hidden layers of three neurodes, 20001 epochs, and 100% training factor for five times.
Here is a partial dataset from five XOR sample runs, as well as the average results and a graph of each scenario's results plotted together.

Epoch	RMSE with five layers					
	1st	2nd	3rd	4th	5th	Avg
0	0.61978726221	0.60575706585	0.61851710988	0.61388101672	0.61380301426	0.61434909378
100	0.50611888526	0.50367127643	0.50444873618	0.50361927609	0.50435451993	0.50444253878
200	0.50214729199	0.50197161066	0.50223191112	0.50310063165	0.50206570583	0.50230343025
300	0.50194164754	0.50190952646	0.50214310319	0.50294432605	0.50202391383	0.50219250341
3100	0.49033415703	0.49755453219	0.49340138555	0.48819756600	0.50146196920	0.49418992199
3200	0.48903430658	0.49701366364	0.49226688387	0.48631446172	0.50143877752	0.49321361867
3300	0.48758849153	0.49640148358	0.49098167883	0.48421751305	0.50141434567	0.49212070253
3400	0.48598368503	0.49570918080	0.48952807557	0.48189565655	0.50138852160	0.49090102391
3500	0.48420690969	0.49492712441	0.48788743301	0.47934190600	0.50136113899	0.48954490242
4000	0.47235576130	0.48927155627	0.47623607996	0.46327576538	0.50119376178	0.48046658494
4100	0.46934529662	0.48769911811	0.47311484979	0.45953406133	0.50115245798	0.47816915677
4200	0.46612267835	0.48594537100	0.46971589597	0.45568528244	0.50110783316	0.47571541218
4300	0.46269803098	0.48399511034	0.46604576669	0.45176336783	0.50105951104	0.47311235738
4400	0.45908665838	0.48183380764	0.46211808196	0.44780014539	0.50100707089	0.47036915285
4500	0.45530850973	0.47944846885	0.45795271049	0.44382320670	0.50095004140	0.46749658743
5000	0.43476889363	0.46394389841	0.43444660602	0.42425685486	0.50057505885	0.45159826235
5100	0.43048577517	0.46015250740	0.42936642266	0.42041844560	0.50047675600	0.44817998137
5200	0.42617441556	0.45616374355	0.42418982902	0.41658108784	0.50036842455	0.44469550010
5300	0.42183296485	0.45199936420	0.41891678162	0.41272521828	0.50024883707	0.44114463320
5400	0.41745037336	0.44768201701	0.41354096839	0.40882789109	0.50011660381	0.43752357073
5500	0.41300593574	0.44323320244	0.40805064918	0.40486337839	0.49997015022	0.43382466319
5600	0.40846911004	0.43867141465	0.40242975839	0.40080371694	0.49980769141	0.43003633829
5700	0.40379961269	0.43401067208	0.39665918634	0.39661924566	0.49962720325	0.42614318400
5800	0.39894786985	0.42925955523	0.39071818729	0.39227919179	0.49942638956	0.42212623874

Epoch	RMSE with five layers					
	1st	2nd	3rd	4th	5th	Avg
7400	0.27234810582	0.33316862424	0.26991015839	0.28812562140	0.48970679749	0.33065186147
7500	0.26368479492	0.32551732944	0.26199427618	0.28020077471	0.48826873583	0.32393318221
7600	0.25536435000	0.31774942230	0.25426726890	0.27237027842	0.48665142765	0.31728054946
7700	0.24740080204	0.30990195570	0.24675554459	0.26467872993	0.48483586288	0.31071457903
7800	0.23979778526	0.30201347302	0.23947935169	0.25716403124	0.48280298004	0.30425152425
7900	0.23255171625	0.29412256231	0.23245310500	0.24985692027	0.48053456396	0.29790377356
8000	0.22565419645	0.28626657784	0.22568593639	0.24278098878	0.47801438722	0.29168041733
8100	0.21909373567	0.27848059160	0.21918237663	0.23595307688	0.47522952950	0.28558786205
8200	0.21285695185	0.27079660364	0.21294308952	0.22938392560	0.47217174002	0.27963046213
8300	0.20692939505	0.26324301157	0.20696559788	0.22307897727	0.46883864258	0.27381112487
15000	0.07695553137	0.07671482085	0.07138560092	0.07922839657	0.12504642378	0.08586615470
15100	0.07634201092	0.07597837951	0.07077327369	0.07854214532	0.12288442330	0.08490404655
15200	0.07574048279	0.07525913075	0.07017380684	0.07786973946	0.12079824358	0.08396828068
15300	0.07515058223	0.07455647475	0.06958679427	0.07721076583	0.11878446583	0.08305781658
15400	0.07457195926	0.07386983865	0.06901184658	0.07656482729	0.11683983116	0.08217166059
15500	0.07400427794	0.07319867506	0.06844859013	0.07593154194	0.11496123560	0.08130886413
15600	0.07344721569	0.07254246079	0.06789666635	0.07531054238	0.11314572475	0.08046852199
15700	0.07290046262	0.07190069547	0.06735573099	0.07470147506	0.11139048807	0.07964977044
15800	0.07236372094	0.07127290044	0.06682545339	0.07410399965	0.10969285317	0.07885178552
15900	0.07183670434	0.07065861759	0.06630551588	0.07351778841	0.10805027983	0.07807378121
16000	0.07131913748	0.07005740828	0.06579561315	0.07294252565	0.10646035414	0.07731500774
16100	0.07081075544	0.06946885240	0.06529545165	0.07237790717	0.10492078266	0.07657474986
16200	0.07031130325	0.06889254736	0.06480474904	0.07182363973	0.10342938666	0.07585232521
16300	0.06982053544	0.06832810726	0.06432323371	0.07127944063	0.10198409655	0.07514708272
16400	0.06933821558	0.06777516204	0.06385064423	0.07074503718	0.10058294638	0.07445840108
16500	0.06886411588	0.06723335669	0.06338672894	0.07022016631	0.09922406862	0.07378568729
19400	0.05790358162	0.05517395452	0.05284164254	0.05822936495	0.07236711585	0.05930313190
19500	0.05760085323	0.05485328806	0.05255548376	0.05790277917	0.07173506845	0.05892949453
19600	0.05730206143	0.05453742799	0.05227332280	0.05758071769	0.07111629875	0.05856196573
19700	0.05700712539	0.05422626250	0.05199507302	0.05726308511	0.07051039046	0.05820038729
19800	0.05671596654	0.05391968326	0.05172065024	0.05694978869	0.06991694410	0.05784460657
19900	0.05642850848	0.05361758527	0.05144997272	0.05664073825	0.06933557620	0.05749447619
20000	0.05614467692	0.05331986672	0.05118296105	0.05633584606	0.06876591854	0.05714985386



According to the provided table and graph, it can be observed that RMSE decreases rapidly during the first 100 epochs in all the five runs. However, after that, the speed of decrease slows down. Although the duration of the initial sharp decrease varies across the five runs, the overall trend remains consistent. There seems to be a significant acceleration in the RMSE decrease after a certain epoch. Notably, the starting times among the five runs vary, with values of 3000, 4000, 4200, 5000, and 8000 from early to late. While the overall trend of the lines is similar, the shape of the lines exhibits greater differences than those observed in the model with fewer layers.

During this particular run, it is evident that the running time is considerably longer than that required for the model with fewer layers. Despite the average RMSE decreasing from 0.6143 to 0.05714, indicating impressive performance, the reduction in error is not as significant as that observed in

the model with three layers. Therefore, it is necessary to evaluate whether the increased computation time is justified by the relatively smaller improvements in performance.

To summarize, while the overall trends of each line in the model with different numbers of layers may be similar, increasing the number of layers leads to a greater diversity in the shape of the lines. Although adding more layers generally results in improved RMSE, the additional computational time required also increases. Furthermore, as the number of layers increases, the improvements in RMSE become progressively smaller. Therefore, finding the optimal number of layers for a given neural network model requires careful consideration of the trade-off between improved performance and increased computational complexity.

Extra Credit Opportunity #3 - Preprocessing Data

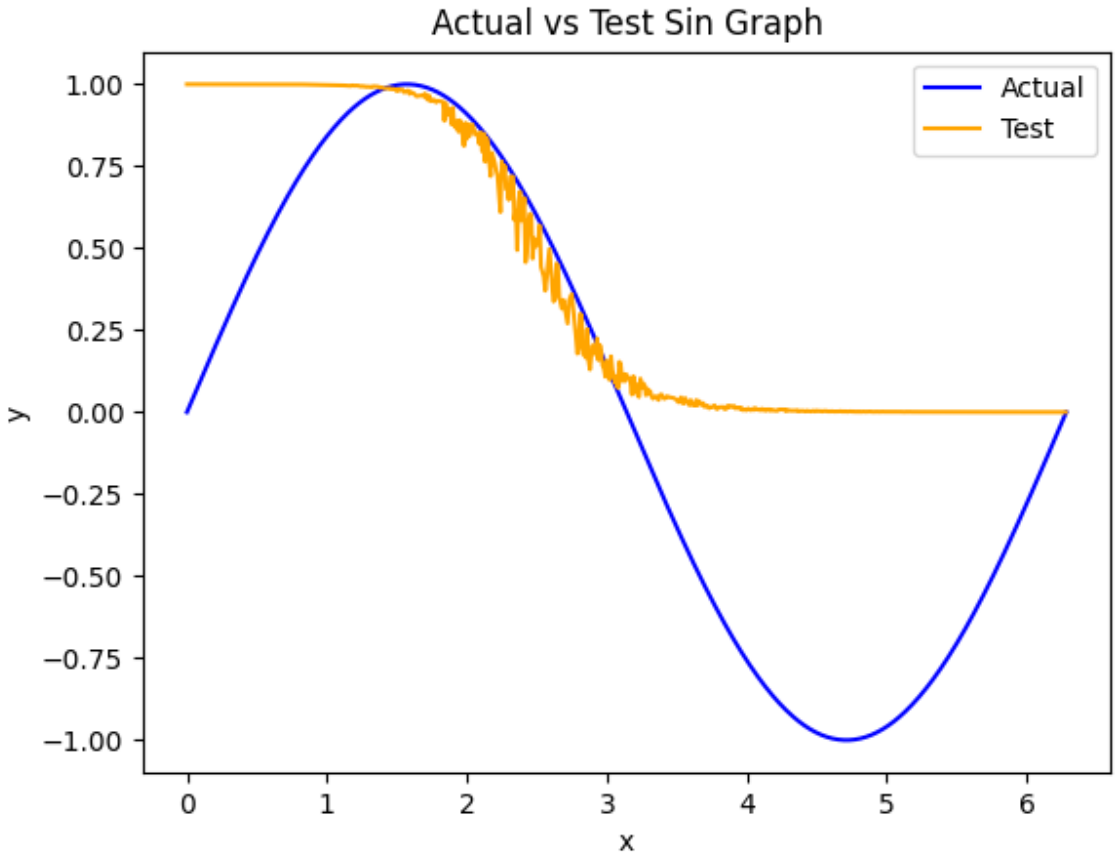
In this section, I trained the sin() function from 0 to 2pi and extending the set, recording the inputs and outputs at the test phase.

Throughout the training process, the initial RMSE was calculated to be 0.969519, and it decreased to 0.519477 by the final epoch, indicating that the model was able to learn and improve over time. However, despite the decrease in RMSE, 0.519477 may still be considered high for a neural network. Moreover, when the model was evaluated on a separate testing dataset, the RMSE was found to be 0.5636, which is relatively high. This indicates that the model may not be generalizing well to new data and may be overfitting to the training dataset. Therefore, further improvements training process may be necessary to improve its performance on the testing dataset.

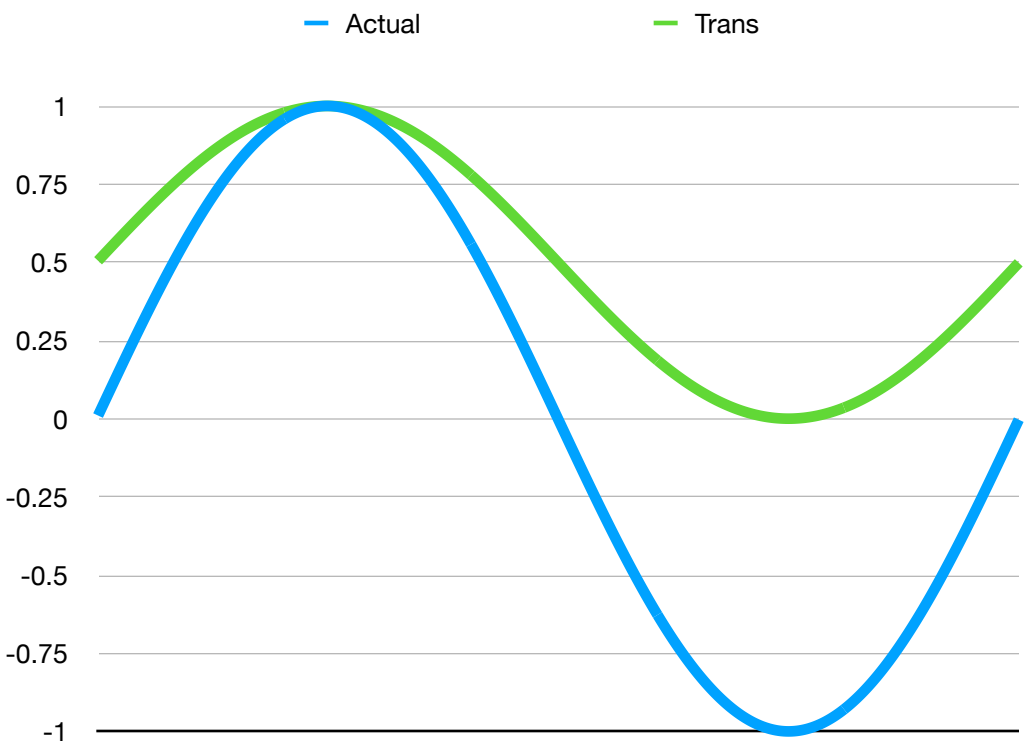
Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.96951895124	2000	0.53807756022	9000	0.52011076373
100	0.65744458844	2100	0.53715102506	9100	0.52004135289
200	0.62407924454	2200	0.53629471319	9200	0.51997338144
300	0.60325975370	2300	0.53550062511	9300	0.51990680714
400	0.58941996266	2400	0.53476195814	9400	0.51984158937
500	0.57957523062	2500	0.53407288767	9500	0.51977768909
600	0.57219955890	2600	0.53342839509	9600	0.51971506871
700	0.56644932209	2700	0.53282413123	9700	0.51965369205
800	0.56182514546	2800	0.53225630683	9800	0.51959352426
900	0.55801427876	2900	0.53172160414	9900	0.51953453179
1000	0.55481127989	3000	0.53121710485	10000	0.51947668227

I plot the actual sin curve, and overlay the test results on the same graph as below. The graph clearly shows a significant difference between the testing results and the actual sine line. The output range of the network is [0,1], whereas the value range of the sin function is [-1,1]. The network did not learn the training data of sin function from 0 to 2pi, possibly because

the model is not designed to handle negative values, as the sin function's value range includes negative values. To address this issue, I will attempt to transform the range to a positive scale of 0 to 1.



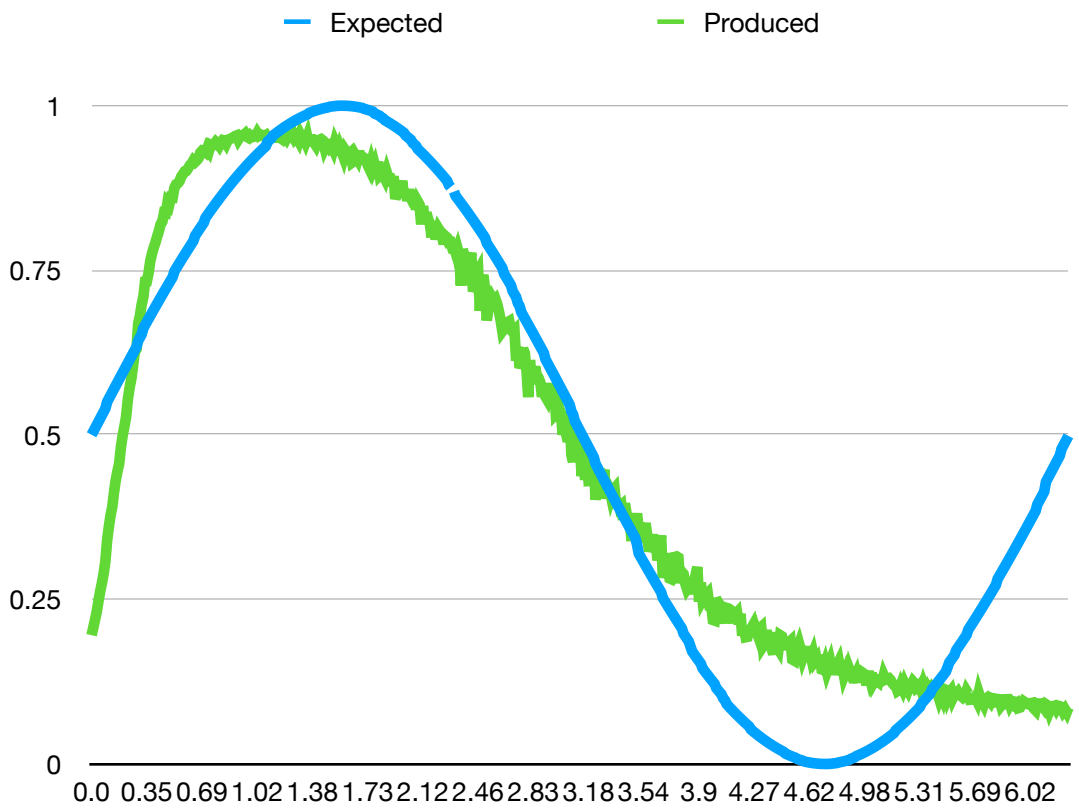
To transform the values of sin, I added 1 to each value and then divided the result by 2. This resulted in a transformed sine function with a range of values between 0 and 1. I plotted both the actual sine function and the transformed sine function together in the graph below.



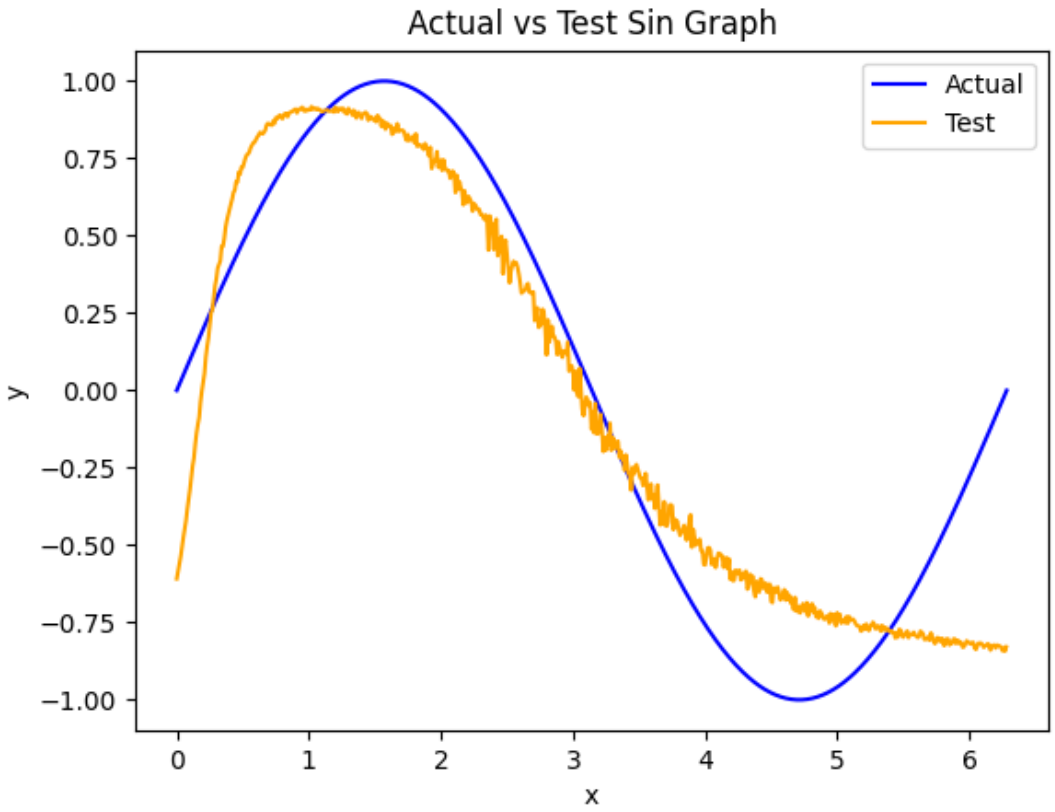
In this training process, the initial RMSE was calculated to be 0.423108. Over the course of training, the RMSE decreased to 0.12806 by the final epoch, indicating that the model was able to learn and improve its performance over time. The testing RMSE was calculated to be 0.12722368. These results suggest that the performance of the training process was significantly improved after transforming the values of sin.

Epoch	RMSE	Epoch	RMSE	Epoch	RMSE	Epoch	RMSE
0	0.42310787034	2000	0.15920444384	6000	0.13378380530	9000	0.12904841174
100	0.33773753557	2100	0.15754284919	6100	0.13355648568	9100	0.12894074502
200	0.30462895165	2200	0.15600822862	6200	0.13333607219	9200	0.12883521025
300	0.27479878973	2300	0.15458781851	6300	0.13312224334	9300	0.12873174021
400	0.25155793902	2400	0.15327041619	6400	0.13291469725	9400	0.12863027052
500	0.23393201937	2500	0.15204605587	6500	0.13271315014	9500	0.12853073949
600	0.22044718722	2600	0.15090582101	6600	0.13251733505	9600	0.12843308798
700	0.20991671873	2700	0.14984173164	6700	0.13232700061	9700	0.12833725928
800	0.20150147452	2800	0.14884666279	6800	0.13214190989	9800	0.12824319896
900	0.19462612787	2900	0.14791427020	6900	0.13196183946	9900	0.12815085478
1000	0.18889521898	3000	0.14703891461	7000	0.13178657839	10000	0.12806017657

I plot a transform sin curve, and overlay the test results on the same graph as below. The graph clearly shows the shape of expected line of produced line is much similar than previous one, indicating the the model is performing better and generating outputs that closely match the expected outcomes.



To transform back the model's output to the original range of [-1,1], I multiplied the output by 2 and then subtracted 1. The graph below shows the actual sine function and the testing results transformed to the range of [-1,1].



The graph shows that the output range is [-1,1], and the transformed data leads to superior results in contrast to the untransformed data. However, certain values of the sine function produce testing outcomes that do not align with the expected results, particularly between 1.5pi and 2pi.