(I) Graph 1:

Test RMSE for a given number of hidden layers vs number of neurons:

1) For all the hidden layers, the pattern is almost similar. The lowest error is for **number of neurons(N)** = 5, and this increases till N=15. After N=15, for N=20, the error decreases, but is still higher than N=10. After this, there is sudden increase in error, and error is highest for N=30.

2) As can be seen from the graphs, the error transition from N=20 to N=30 is steep.

3) The large error in N=30 might be due to overfitting. RMSE is between 9.48 to 9.68. 9.48 is error for N=5 neurons, 9.49 for N=10, 9.53 for N=15, 9.495 for N=20 and 9.68 for N=30.

(II) Graph 2:

Test RMSE for a given number of neurons vs number of hidden layers:

1) This corresponds to the observations we made in Graph 1. In general, 9.48 is error for N=5 neurons, 9.49 for N=10, 9.53 for N=15, 9.495 for N=20 and 9.68 for N=30.

2) But for a given N, across all the **number of hidden layers L**; error till 2 digits after the decimal place, is same.

3) For all reasons, we can assume RMSE to be constant for a given N, and varying hidden layer number L. Still, for N=5, RMSE is lowest for L=1; for N=10, RMSE is lowest for L=8; for N=15, RMSE is lowest for L=4; for N=20, RMSE is lowest for L=4; and for N=30, RMSE is lowest for L=8.

(III) Graph 3:

Test RMSE for a given number of hidden layers(L) vs Average epoch time:

1) From this, we can see that for L=1, when average epoch time is low, RMSE is also low. As we increase epoch time, the RMSE also increases. At one point, the average epoch time, increases, but RMSE decreases. This might be because model increased in complexity, but overfitting has not taken place yet. But afterwards, when epoch time is increased further, RMSE shoots up, showing overfitting. From values, we can see that overfitting RMSE if for N=30.

2) For L=2, the overfitting case is in fact, taking lesser epoch time than an N which is giving lesser RMSE. N=20 is giving much less RMSE than R=30, but is taking more average epoch time.

3) For L=4, the overfitting case(high RMSE) is again taking the most epoch time.Again, this follows the same pattern as 1), where RMSE was increasing with epoch time, but suddenly fell for a model(N=20). For N=30, it still has the highest epoch time and RMSE.

4) For L=6, again same pattern as 1),3) follows. You can see that the average epoch time for N=30 is in fact equal, and even at times, lower than N=20. The values of the epoch times confirm this. Thats why the line of the high RMSE N=30 case, bends a bit backwards to account for low epoch time.

5) For L=8, again same pattern as 1),3),4) follows. But here, N=30 is definitely taking more average epoch time than N=20. So, line is forwards. N=5 is always taking least epoch time.

(IV) Graph 4:

CPU/GPU/TPU total program execution time for 5 neurons vs No of layers(L):

1) CPU time is highest for L=2 layers, and decreases progressively as L=8, 4, 6, 1. So, there is a steep increase from L=1 to L=2. After this, there is a mild decrease till L=6, after which CPU time increases a bit again.

2) GPU time is lowest for L=2, and highest for L=1.There is a steep decrease from L=1 to L=1, after which there is a slow increase till L=8. So, this is in many ways opposite to CPU time: here time is least for L=2, whereas there, it had been highest; here there is a steep decrease from L=1 to L=2, whereas there, it had been an increase; Here, there is an increase from L=2 to L=6, whereas, there, it had been a decrease.

3) TPU time is least for L=1, and it increases for L=2. So, this is same as CPU time. But from L=2 to L=6, TPU time increases, so this is like GPU. TPU time is highest for L=6. There is a decrease in execution time from L=6 to L=8, which did not happen in CPU or GPU.

4) In total, there is an increase in execution time as L increases. This is expected.

(V) Graph 5:

CPU time vs GPU time, CPU time vs TPU time, GPU time vs TPU time:

1) When CPU time is decreasing, GPU time is increasing. But there are instances after CPU time 80 sec, that, for the same CPU time, more than 1 GPU time is existing; and for the same GPU time, more than 1 CPU times can correspond. CPU time increases from 40 to 120, and GPU time correspondingly decreases from 180 to 120. Thus when CPU time taken is highest, that GPU time taken is lowest and almost equal to it.

2) At its lowest, CPU time and TPU time are at same ~ 60 seconds. From there, CPU time increases to ~130 seconds, whereas TPU time doesnt cross 100. Thus, increase in TPU time with layers is not much. There is still, a direct correspondence between CPU and TPU times.

3) GPU time starts from above 120, and crosses 170, whereas TPU time doesnt cross 100. Thus, there is a decrease in TPU time as GPU time increases.
In small neural networks such as this, with 5 neurons in each hidden layer, CPU performs better than GPU in general, and GPU is better in case of networks with large number of parameters.

(VI) Graph 6:

Average CPU, GPU, TPU epoch times comparison:

1) Avg epoch time for CPU crosses 20 only once, while for GPU, it is quite high.

2) Average CPU time is highest for L=2 (~21). Lowest is for L=1(~13). There is a decrease after L=2, and then, an increase.This is same as the CPU total execution time versus number of layers.

3) Average GPU epoch time is highest for L=1(~32) and lowest for L=2(~23). After this, there is an increase. This is again, similar to the graph we saw of Total execution time vs number of layers.

4) Average TPU time constantly increases for L=1(~13) to L=8(~15). Thus, TPU performs best, in all, and is least sensitive to change in number of layers.

(VII) Graph 7:

Epoch time vs Number of neurons:

1) For N=5, epoch time increases linearly.There is a decrease from N=20 to N=30 though.

2) For N=10, minimum epoch time is at N=10. Maximumis at N=5.

3) For N=15, there is a constant increase of epoch time with number of neurons.

4) For N=20, there is an increase till N=20, after which N=30 time is less. So, in quite a few of the cases, we saw that average epoch tie for N=20 was more for N=20 than N=30.

(VIII) Graph 8:

Epoch time vs Number of Layers:

1) The pattern for N=5 is familiar, where, max time is for L=2 and min for L=1, and after L=2, there is an increase.

2) In fact, in all the other N=10, 15, 20, 30; there is a steady increase of epoch time with number of layers. The reason I chose N=5 for CPU, GPU, TPU analysis was because its RMSE was minimum and I thought of picking the best.
But if I had chosen any of the other N's, maybe I would have gotten a lot linear relationships.

NOTES:

1) The dimensions of each layer, and number of neurons used is explained in the program. Details of how the program works is also mentioned.

The RMSE for a fixed number of neurons, across different number of hidden layers is:

```
no_layers=[1,2,4,6,8]
```

```
test_rmse_5_neurons=[9.484940624412207,9.485241945286925,9.4852419389164,9.485902832832021,9.485242046758994] # Test_RMSE for 5 neurons in hidden layer. Every RMSE is for no. of hidden layers= 1, 2, 4, 6, 8 respectively.
test_rmse_10_neurons=[9.490749679947562,9.490746598472864,9.490748778614062,9.490749809794034,9.490714140099996]
test_rmse_15_neurons=[9.534213482744976,9.534189461563852,9.534178210583951,9.534217198065976,9.534182339295494]
test_rmse_20_neurons=[9.494937858187168,9.494938629272601,9.494936070057676,9.494940927079748,9.494934737169235]
test_rmse_30_neurons=[9.684310754776556,9.684311832759356,9.684311811300702,9.684310771895532,9.684310697014213]
```

2) Since the minimum RMSE for me was for N=5, I have kept N=5 as constant and done analysis of CPU, GPU, TPU by varying the number of layers.

Epoch times for normal CPU are:

```
no_neurons=[5,10,15,20,30]

# For no. of hidden layers=1
epoch_times_5_neurons_1_layer=[14.871137651000026, 13.570384841001214, 13.88249884600009] # I used early stopping in my program. So, only epoch times of those iterations are there, which gave a better RMSE than the previous one.
epoch_times_10_neurons_1_layer=[14.685702896000294, 13.824583472000086, 13.550479282999731, 13.752716221999435]
epoch_times_15_neurons_1_layer=[16.119246821999695, 15.360740249001537, 15.284523345999332, 15.350034574999881]
epoch_times_20_neurons_1_layer=[17.868733545001305, 16.198592030999862, 16.308716334002384, 16.259773826997844]
epoch_times_30_neurons_1_layer=[17.32878567700027, 16.060801800998888, 16.172498924999672, 16.114753481000662, 16.416920325998944]

# For no. of hidden layers=2
epoch_times_5_neurons_2_layer=[21.697007921000477, 20.93782656300027, 20.91623797200009, 20.760044058999483, 21.89751109200006, 21.76080045499839]
epoch_times_10_neurons_2_layer=[14.965459798997472, 14.385745997999038, 13.956001604998164, 14.06257141900278]
epoch_times_15_neurons_2_layer=[18.671506812999723, 16.603440939001302, 16.966852713001572, 16.77439511099874]
epoch_times_20_neurons_2_layer=[18.071632928997133, 18.73847734899755, 16.630607307997707, 17.02174010499948]
epoch_times_30_neurons_2_layer=[17.69449215699933, 17.14842911699816, 16.988785511999595, 16.810753286001272, 16.945490025002073]

# For no. of hidden layers=4
epoch_times_5_neurons_4_layer=[13.373188697158605, 9.703955928009178, 9.702230603159084, 9.70211630190503, 9.702114409491298]
epoch_times_10_neurons_4_layer=[17.491342206998524, 14.988746151997475,
```

```
14.973542265001015, 14.820126185000845]
epoch_times_15_neurons_4_layer=[18.168999838999298, 17.3939374770016,
17.21373718399991, 17.33145760199841] #This means, for 4 hidden layers with 15
neurons EACH, the number of elements in list is the number of epochs that took
place, and epoch time is the entry in the list.
epoch_times_20_neurons_4_layer=[18.884539616003167, 18.373712567998155,
18.338266222999664, 18.251989608001168]
epoch_times_30_neurons_4_layer=[18.86760795200098, 17.947390000001178,
19.648128980003094, 18.522724637001375, 17.99134933499954]

# For no. of hidden layers=6
epoch_times_5_neurons_6_layer= [16.137721058999887, 16.31683720699948,
15.107018773000163, 15.034416644000885, 15.131612571000005]
epoch_times_10_neurons_6_layer=[17.07463467100024, 16.008795460002148,
15.680057773999579, 15.831174379000004]
epoch_times_15_neurons_6_layer=[19.179884443001356, 18.779515409998567,
18.64876903699769, 19.421667107002577]
epoch_times_20_neurons_6_layer=[20.113473908000742, 19.56500863799738,
19.370590705999348, 22.231488797002385]
epoch_times_30_neurons_6_layer=[19.869464693001646, 18.99948091999977,
19.60865139299858, 19.292971265000233, 19.116937839997263]

# For no. of hidden layers=8
epoch_times_5_neurons_8_layer= [16.489735160001146, 15.804078242999822,
16.092863080000825, 16.147970229001658, 15.460394667999935,
15.9131394579999321
epoch_times_10_neurons_8_layer=[17.05362054700163, 16.175794117996702,
16.54727813000136, 16.18309016800049]
epoch_times_15_neurons_8_layer=[20.086583292002615, 20.26363766400027,
19.93664997900123, 19.666001457000675]
epoch_times_20_neurons_8_layer=[22.14329101399926, 20.464176825000322,
20.46212871899843, 20.482320358001743]
epoch_times_30_neurons_8_layer= [26.200624151002557, 20.902267943998595,
20.3123158840026, 20.517774078998627, 20.66734335699948]
```