REPORT

The following are some observations related to the binary operation XOR when the all the combinations of changing the number of nodes and changing the number of training samples was tested. The training set was generated by adding Gaussian noise with mean 0 and standard deviation 0.1 to all the combinations (0,0), (1,1), (1,0), (0,1). The trained model was tested on the samples (0,0), (1,1), (1,0), (0,1) (without adding any noise).

The model was trained for 100 epochs.

1.) More samples for a given number of nodes:
Two examples were considered. Both of them had the same learning rate(=0.1) and the same number of hidden nodes(=5). The results are as follows:

For 20 training samples, three out of four predictions were correct with a very less margin from the threshold. Even very less noise in test samples might affect the prediction. The final loss was around 0.14.
For 40 training samples, the performance was better compared to the previous case. All the test samples were predicted correctly and with a much better margin from the threshold (approximately 0.3). The final loss was around 0.001.

So, the number of meaningful training samples will directly affect the performance of the model for a fixed architecture.
(Note: We might not get the same loss value when the model is retrained. This is due to the noise added in the beginning. However, the margins are almost the same when the model was retrained and tested.)

2.) More nodes for a given number of samples:
Three examples were considered. All of them had the same learning rate(=0.1) and the same number of samples(=40). The results are as follows:

For 5 hidden nodes, all predictions were correct but with a less margin from the threshold(up to 0.2). Also, the final loss was around 0.18
For 10 hidden nodes, all predictions were correct and with a higher margin from the threshold(0.3-0.4). The final loss was around 0.01
For 50 hidden nodes, two out of four predictions were correct. Although looking at the output sigmoid value, it can be observed that there was a clear overfit(with an inclination to one of the classes)
So, more nodes do improve the model performance. However, arbitrarily large number of nodes for a limited training data leads to models which overfit.

3.) Learning Rate:
The choice of learning rate also affects the performance of the model. Very low learning rates lead to slow convergence of model whereas very high lead to missing of the optimal minimum value. A learning rate of 0.1 was found to perform well on this particular problem. For a learning rate of 1 the performance was very bad and there was a slow convergence for learning rate of 0.01. The model consists of 10 hidden nodes with 40 training samples.