F1 score is interpreted as a weighted average of the precision and recall (termed as the harmonic mean of precision and recall). **It is a measure of a test's accuracy**. A better F1 score meaning a better accuracy of the model.

The above graph is achieved by training the model for 30 max epochs with a learning rate of 0.01 (10 hidden units for NN model). For each e from 1 to max epoch, the network is trained for e epochs. From the above graph we have below observations:

1. As the model is trained using training dataset, it is expected that plot for train dataset shows better F1 score than test dataset. This we can see happening after epoch 12+. This shows that the network is overfitting for the training dataset after epoch 12+. But until below that epoch, we see that test F1 score is outperforming the training dataset. So, **we can say that the network is trained very well to predict on an unknown dataset with better accuracy/F1 score until epoch <12 after which the network is getting overfitted to train dataset hence performance on train dataset overtakes test dataset.**
2. Comparing two different graphs above, the Neural Network is performing better than the logistic network. This is because of the hidden units which bring in more scope for training the network.