Homework 4

As instructed, I cloned the Kaldi from it's repo and installed it on my Ubuntu OS.

Issues Faced

1. https://github.com/kaldi-asr/kaldi/blob/master/egs/timit/s5/run.sh. On line 144 there is a random exit 0 which exits the code unexpectedly
2. I helped in solving the following bug in their official code - Since TIMIT is a small dataset I was getting an error (reduce nj). But instead we had to change the variable train_nj. Changing the variable nj did not solve the bug. https://github.com/kaldi-asr/kaldi/issues/3716

The columns represent the following-

1. Word Error Rate x 100
2. Number of Sentences
3. Number of Words
4. Correctly Identified Words
5. Substitution
6. Deletion
7. Insertion
8. Word Error Rate
9. Sentence Error Rate
10. NCE - Quality of Confidence Score

Comparing all ASR model based on WER score. Lower the WER score better the model. We are comparing the following models

1. Monophone Delta
2. Tri1 - first triphone system (delta+delta-delta features)
3. Tri2 : an LDA+MLLT system
4. Tri3 : Speaker Adaptive Training (SAT) system
5. SGMM2
6. SGMM2 + MMI Training
7. Tri4_nnet - DNN Hybrid System

On Dev Set:

1. SGMM2 and SGMM2+MMI training models have the best WER score (~18). Both models have correctly identified ~84% of total words.
2. DNN based model has the second best WER (20.5). It has correctly identified ~82.3% of total words.
3. The monophone model has the worst performance. It's WER=31.7 and identifies 72.1% of total words correctly.
4. Therefore, SGMM2 is the best and Monophone Model is the worst.

On Test Set:

1. The analysis from the top is similar to the validation phase.
2. SGMM2 and SGMM2 + MMI models have equal performance (WER = 19.3). They correctly identified ~83% of total words.
3. The DNN model's performance is the second best (WER~22.1) and ~81.5% of total words correctly.
4. Insertion Score is best in SGMM2 (2.3) while Deletion score is best in SAT system (4.3)
5. The monophone model has the worst WER score (32.4).

Conclusion:

SGMM2 and SGMM2+MMI are the best models. DNN Hybid is the 2nd best model. And the monophone model is the worst

Results For Hidden Layer 2

```
================================================================================
                    Getting Results [see RESULTS file]
================================================================================
%WER 31.7 | 400 15057 | 71.7 19.4 8.8 3.4 31.7 100.0 | -0.471 | exp/mono/decode_dev/score_5/ctm_39phn.filt.sys
%WER 24.9 | 400 15057 | 79.1 15.8 5.2 3.9 24.9 99.8 | -0.128 | exp/tri1/decode_dev/score_10/ctm_39phn.filt.sys
%WER 23.3 | 400 15057 | 80.5 14.6 5.0 3.8 23.3 99.8 | -0.280 | exp/tri2/decode_dev/score_10/ctm_39phn.filt.sys
%WER 20.5 | 400 15057 | 83.0 12.7 4.3 3.5 20.5 99.8 | -0.859 | exp/tri3/decode_dev/score_8/ctm_39phn.filt.sys
%WER 23.4 | 400 15057 | 79.9 14.8 5.3 3.4 23.4 99.8 | -0.176 | exp/tri3/decode_dev.si/score_10/ctm_39phn.filt.sys
%WER 20.5 | 400 15057 | 82.3 11.9 5.7 2.9 20.5 100.0 | -0.406 | exp/tri4_nnet/decode_dev/score_6/ctm_39phn.filt.sys
%WER 17.9 | 400 15057 | 84.3 11.1 4.6 2.2 17.9 99.5 | -0.165 | exp/sgmm2_4/decode_dev/score_10/ctm_39phn.filt.sys
%WER 18.1 | 400 15057 | 84.4 11.3 4.3 2.6 18.1 99.5 | -0.157 | exp/sgmm2_4_mmi_b0.1/decode_dev_it1/score_10/ctm_39phn.filt.sys
%WER 18.0 | 400 15057 | 85.1 11.1 3.8 3.1 18.0 99.3 | -0.326 | exp/sgmm2_4_mmi_b0.1/decode_dev_it2/score_8/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.5 11.2 4.3 2.7 18.2 99.5 | -0.180 | exp/sgmm2_4_mmi_b0.1/decode_dev_it3/score_10/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.7 11.3 4.0 2.9 18.2 99.3 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_dev_it4/score_9/ctm_39phn.filt.sys
%WER 32.2 | 192 7215 | 71.2 19.1 9.6 3.4 32.2 100.0 | -0.460 | exp/mono/decode_test/score_5/ctm_39phn.filt.sys
%WER 25.8 | 192 7215 | 77.8 16.8 5.4 3.6 25.8 100.0 | -0.128 | exp/tri1/decode_test/score_10/ctm_39phn.filt.sys
%WER 23.7 | 192 7215 | 79.8 14.9 5.3 3.5 23.7 99.5 | -0.259 | exp/tri2/decode_test/score_10/ctm_39phn.filt.sys
%WER 21.2 | 192 7215 | 82.2 13.5 4.3 3.4 21.2 99.0 | -0.855 | exp/tri3/decode_test/score_8/ctm_39phn.filt.sys
%WER 23.9 | 192 7215 | 79.3 15.0 5.6 3.3 23.9 99.0 | -0.219 | exp/tri3/decode_test.si/score_10/ctm_39phn.filt.sys
%WER 22.1 | 192 7215 | 81.5 13.1 5.3 3.7 22.1 100.0 | -0.964 | exp/tri4_nnet/decode_test/score_4/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.0 12.1 4.9 2.3 19.3 100.0 | -0.155 | exp/sgmm2_4/decode_test/score_10/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.6 12.0 4.4 2.9 19.3 100.0 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_test_it1/score_9/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.0 12.2 3.8 3.4 19.4 100.0 | -0.480 | exp/sgmm2_4_mmi_b0.1/decode_test_it2/score_7/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 83.9 12.0 4.0 3.3 19.4 100.0 | -0.386 | exp/sgmm2_4_mmi_b0.1/decode_test_it3/score_8/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.1 12.1 3.8 3.5 19.4 100.0 | -0.528 | exp/sgmm2_4_mmi_b0.1/decode_test_it4/score_7/ctm_39phn.filt.sys
================================================================================
Finished successfully on Fri Nov 15 22:12:11 EST 2019
================================================================================
rushabh@rushabh-G7-7588:~/Desktop/dlsp/hw4/kaldi/egs/timit/s5$
```

Comparison of results of DNN models with different numbers of hidden layers:
The results show that as the number of hidden layers increase the ASR model gets better But the decrease in WER is very low (ranges from 20.8 to 20.1 in dev set and 22.2 to 21.9 in the test set) .

Validation set:

1. For 1 hidden layer, 82.4% of the words are correctly identified and WER 20.8.
2. For 2 hidden layers, 82.3% of the words are correctly identified and WER 20.5 slightly better than 1.
3. For 3 hidden layers, 82.7% of the words are correctly identified and WER 20.4 slightly better than 2.
4. For 4 hidden layers, 83% of the words are correctly identified and WER 20.1 slightly better than 3.

Testing set:

1. For 1 hidden layer, 81.1% of the words are correctly identified and WER 22.2.
2. For 2 hidden layers, 81.5% of the words are correctly identified and WER 22.1 slightly better than 1.
3. For 3 hidden layers, 81.3% of the words are correctly identified and WER 21.9 slightly better than 2.
4. For 4 hidden layers, 80.9% of the words are correctly identified and WER 21.9 similar to 3.
5. We can notice that even though the WER is similar in models with 3 and 4 hidden layers the percentage of correctly identified words is less in the model with 4 hidden layers (80.9%) in comparison to model with 3 hidden layers (81.3%). So there could be a chance that the model is overfitting as the number of deletions are increasing (5.8 in model with 3 hidden layers v/s 6.3 in model with 4 hidden layers) .

Results For Hidden Layer 1

```
================================================================================
                    Getting Results [see RESULTS file]
================================================================================
%WER 31.7 | 400 15057 | 71.7 19.4 8.8 3.4 31.7 100.0 | -0.471 | exp/mono/decode_dev/score_5/ctm_39phn.filt.sys
%WER 24.9 | 400 15057 | 79.1 15.8 5.2 3.9 24.9 99.8 | -0.128 | exp/tri1/decode_dev/score_10/ctm_39phn.filt.sys
%WER 23.3 | 400 15057 | 80.5 14.6 5.0 3.8 23.3 99.8 | -0.280 | exp/tri2/decode_dev/score_10/ctm_39phn.filt.sys
%WER 20.5 | 400 15057 | 83.0 12.7 4.3 3.5 20.5 99.8 | -0.859 | exp/tri3/decode_dev/score_8/ctm_39phn.filt.sys
%WER 23.4 | 400 15057 | 79.9 14.8 5.3 3.4 23.4 99.8 | -0.176 | exp/tri3/decode_dev.si/score_10/ctm_39phn.filt.sys
%WER 20.8 | 400 15057 | 82.4 12.2 5.5 3.2 20.8 99.8 | -0.599 | exp/tri4_nnet/decode_dev/score_5/ctm_39phn.filt.sys
%WER 17.9 | 400 15057 | 84.3 11.1 4.6 2.2 17.9 99.5 | -0.165 | exp/sgmm2_4/decode_dev/score_10/ctm_39phn.filt.sys
%WER 18.1 | 400 15057 | 84.4 11.3 4.3 2.6 18.1 99.5 | -0.157 | exp/sgmm2_4_mmi_b0.1/decode_dev_it1/score_10/ctm_39phn.filt.sys
%WER 18.0 | 400 15057 | 85.1 11.1 3.8 3.1 18.0 99.3 | -0.326 | exp/sgmm2_4_mmi_b0.1/decode_dev_it2/score_8/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.5 11.2 4.3 2.7 18.2 99.5 | -0.180 | exp/sgmm2_4_mmi_b0.1/decode_dev_it3/score_10/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.7 11.3 4.0 2.9 18.2 99.3 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_dev_it4/score_9/ctm_39phn.filt.sys
%WER 32.2 | 192 7215 | 71.2 19.1 9.6 3.4 32.2 100.0 | -0.460 | exp/mono/decode_test/score_5/ctm_39phn.filt.sys
%WER 25.8 | 192 7215 | 77.8 16.8 5.4 3.6 25.8 100.0 | -0.128 | exp/tri1/decode_test/score_10/ctm_39phn.filt.sys
%WER 23.7 | 192 7215 | 79.8 14.9 5.3 3.5 23.7 99.5 | -0.259 | exp/tri2/decode_test/score_10/ctm_39phn.filt.sys
%WER 21.2 | 192 7215 | 82.2 13.5 4.3 3.4 21.2 99.0 | -0.855 | exp/tri3/decode_test/score_8/ctm_39phn.filt.sys
%WER 23.9 | 192 7215 | 79.3 15.0 5.6 3.3 23.9 99.0 | -0.219 | exp/tri3/decode_test.si/score_10/ctm_39phn.filt.sys
%WER 22.2 | 192 7215 | 81.1 13.3 5.6 3.3 22.2 99.5 | -0.965 | exp/tri4_nnet/decode_test/score_4/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.0 12.1 4.9 2.3 19.3 100.0 | -0.155 | exp/sgmm2_4/decode_test/score_10/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.6 12.0 4.4 2.9 19.3 100.0 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_test_it1/score_9/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.0 12.2 3.8 3.4 19.4 100.0 | -0.480 | exp/sgmm2_4_mmi_b0.1/decode_test_it2/score_7/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 83.9 12.0 4.0 3.3 19.4 100.0 | -0.386 | exp/sgmm2_4_mmi_b0.1/decode_test_it3/score_8/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.1 12.1 3.8 3.5 19.4 100.0 | -0.528 | exp/sgmm2_4_mmi_b0.1/decode_test_it4/score_7/ctm_39phn.filt.sys
================================================================================
Finished successfully on Fri Nov 15 23:25:42 EST 2019
================================================================================
```

Results For Hidden Layer 3

```
====================================================================
                  Getting Results [see RESULTS file]
====================================================================
%WER 31.7 | 400 15057 | 71.7 19.4 8.8 3.4 31.7 100.0 | -0.471 | exp/mono/decode_dev/score_5/ctm_39phn.filt.sys
%WER 24.9 | 400 15057 | 79.1 15.8 5.2 3.9 24.9 99.8 | -0.128 | exp/tri1/decode_dev/score_10/ctm_39phn.filt.sys
%WER 23.3 | 400 15057 | 80.5 14.6 5.0 3.8 23.3 99.8 | -0.280 | exp/tri2/decode_dev/score_10/ctm_39phn.filt.sys
%WER 20.5 | 400 15057 | 83.0 12.7 4.3 3.5 20.5 99.8 | -0.859 | exp/tri3/decode_dev/score_8/ctm_39phn.filt.sys
%WER 23.4 | 400 15057 | 79.9 14.8 5.3 3.4 23.4 99.8 | -0.176 | exp/tri3/decode_dev.si/score_10/ctm_39phn.filt.sys
%WER 20.4 | 400 15057 | 82.7 12.0 5.3 3.2 20.4 99.8 | -0.632 | exp/tri4_nnet/decode_dev/score_5/ctm_39phn.filt.sys
%WER 17.9 | 400 15057 | 84.3 11.1 4.6 2.2 17.9 99.5 | -0.165 | exp/sgmm2_4/decode_dev/score_10/ctm_39phn.filt.sys
%WER 18.1 | 400 15057 | 84.4 11.3 4.3 2.6 18.1 99.5 | -0.157 | exp/sgmm2_4_mmi_b0.1/decode_dev_it1/score_10/ctm_39phn.filt.sys
%WER 18.0 | 400 15057 | 85.1 11.1 3.8 3.1 18.0 99.3 | -0.326 | exp/sgmm2_4_mmi_b0.1/decode_dev_it2/score_8/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.5 11.2 4.3 2.7 18.2 99.5 | -0.180 | exp/sgmm2_4_mmi_b0.1/decode_dev_it3/score_10/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.7 11.3 4.0 2.9 18.2 99.3 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_dev_it4/score_9/ctm_39phn.filt.sys
%WER 32.2 | 192 7215 | 71.2 19.1 9.6 3.4 32.2 100.0 | -0.460 | exp/mono/decode_test/score_5/ctm_39phn.filt.sys
%WER 25.8 | 192 7215 | 77.8 16.8 5.4 3.6 25.8 100.0 | -0.128 | exp/tri1/decode_test/score_10/ctm_39phn.filt.sys
%WER 23.7 | 192 7215 | 79.8 14.9 5.3 3.5 23.7 99.5 | -0.259 | exp/tri2/decode_test/score_10/ctm_39phn.filt.sys
%WER 21.2 | 192 7215 | 82.2 13.5 4.3 3.4 21.2 99.0 | -0.855 | exp/tri3/decode_test/score_8/ctm_39phn.filt.sys
%WER 23.9 | 192 7215 | 79.3 15.0 5.6 3.3 23.9 99.0 | -0.219 | exp/tri3/decode_test.si/score_10/ctm_39phn.filt.sys
%WER 21.9 | 192 7215 | 81.3 12.8 5.8 3.2 21.9 99.5 | -0.690 | exp/tri4_nnet/decode_test/score_5/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.0 12.1 4.9 2.3 19.3 100.0 | -0.155 | exp/sgmm2_4/decode_test/score_10/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.6 12.0 4.4 2.9 19.3 100.0 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_test_it1/score_9/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.0 12.2 3.8 3.4 19.4 100.0 | -0.480 | exp/sgmm2_4_mmi_b0.1/decode_test_it2/score_7/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 83.9 12.0 4.0 3.3 19.4 100.0 | -0.386 | exp/sgmm2_4_mmi_b0.1/decode_test_it3/score_8/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.1 12.1 3.8 3.5 19.4 100.0 | -0.528 | exp/sgmm2_4_mmi_b0.1/decode_test_it4/score_7/ctm_39phn.filt.sys
====================================================================
Finished successfully on Wed Nov 20 01:57:44 EST 2019
====================================================================
(base) rushabh@rushabh-G7-7588:~/Desktop/dlsp/hw4/kaldi/egs/timit/s5$
```

Results For Hidden Layer 4

```
====================================================================
                  Getting Results [see RESULTS file]
====================================================================
%WER 31.7 | 400 15057 | 71.7 19.4 8.8 3.4 31.7 100.0 | -0.471 | exp/mono/decode_dev/score_5/ctm_39phn.filt.sys
%WER 24.9 | 400 15057 | 79.1 15.8 5.2 3.9 24.9 99.8 | -0.128 | exp/tri1/decode_dev/score_10/ctm_39phn.filt.sys
%WER 23.3 | 400 15057 | 80.5 14.6 5.0 3.8 23.3 99.8 | -0.280 | exp/tri2/decode_dev/score_10/ctm_39phn.filt.sys
%WER 20.5 | 400 15057 | 83.0 12.7 4.3 3.5 20.5 99.8 | -0.859 | exp/tri3/decode_dev/score_8/ctm_39phn.filt.sys
%WER 23.4 | 400 15057 | 79.9 14.8 5.3 3.4 23.4 99.8 | -0.176 | exp/tri3/decode_dev.si/score_10/ctm_39phn.filt.sys
%WER 20.1 | 400 15057 | 83.0 11.7 5.3 3.1 20.1 99.8 | -0.570 | exp/tri4_nnet/decode_dev/score_5/ctm_39phn.filt.sys
%WER 17.9 | 400 15057 | 84.3 11.1 4.6 2.2 17.9 99.5 | -0.165 | exp/sgmm2_4/decode_dev/score_10/ctm_39phn.filt.sys
%WER 18.1 | 400 15057 | 84.4 11.3 4.3 2.6 18.1 99.5 | -0.157 | exp/sgmm2_4_mmi_b0.1/decode_dev_it1/score_10/ctm_39phn.filt.sys
%WER 18.0 | 400 15057 | 85.1 11.1 3.8 3.1 18.0 99.3 | -0.326 | exp/sgmm2_4_mmi_b0.1/decode_dev_it2/score_8/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.5 11.2 4.3 2.7 18.2 99.5 | -0.180 | exp/sgmm2_4_mmi_b0.1/decode_dev_it3/score_10/ctm_39phn.filt.sys
%WER 18.2 | 400 15057 | 84.7 11.3 4.0 2.9 18.2 99.3 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_dev_it4/score_9/ctm_39phn.filt.sys
%WER 32.2 | 192 7215 | 71.2 19.1 9.6 3.4 32.2 100.0 | -0.460 | exp/mono/decode_test/score_5/ctm_39phn.filt.sys
%WER 25.8 | 192 7215 | 77.8 16.8 5.4 3.6 25.8 100.0 | -0.128 | exp/tri1/decode_test/score_10/ctm_39phn.filt.sys
%WER 23.7 | 192 7215 | 79.8 14.9 5.3 3.5 23.7 99.5 | -0.259 | exp/tri2/decode_test/score_10/ctm_39phn.filt.sys
%WER 21.2 | 192 7215 | 82.2 13.5 4.3 3.4 21.2 99.0 | -0.855 | exp/tri3/decode_test/score_8/ctm_39phn.filt.sys
%WER 23.9 | 192 7215 | 79.3 15.0 5.6 3.3 23.9 99.0 | -0.219 | exp/tri3/decode_test.si/score_10/ctm_39phn.filt.sys
%WER 21.9 | 192 7215 | 80.9 12.8 6.3 2.7 21.9 99.5 | -0.345 | exp/tri4_nnet/decode_test/score_6/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.0 12.1 4.9 2.3 19.3 100.0 | -0.155 | exp/sgmm2_4/decode_test/score_10/ctm_39phn.filt.sys
%WER 19.3 | 192 7215 | 83.6 12.0 4.4 2.9 19.3 100.0 | -0.249 | exp/sgmm2_4_mmi_b0.1/decode_test_it1/score_9/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.0 12.2 3.8 3.4 19.4 100.0 | -0.480 | exp/sgmm2_4_mmi_b0.1/decode_test_it2/score_7/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 83.9 12.0 4.0 3.3 19.4 100.0 | -0.386 | exp/sgmm2_4_mmi_b0.1/decode_test_it3/score_8/ctm_39phn.filt.sys
%WER 19.4 | 192 7215 | 84.1 12.1 3.8 3.5 19.4 100.0 | -0.528 | exp/sgmm2_4_mmi_b0.1/decode_test_it4/score_7/ctm_39phn.filt.sys
====================================================================
Finished successfully on Wed Nov 20 00:44:32 EST 2019
====================================================================
(base) rushabh@rushabh-G7-7588:~/Desktop/dlsp/hw4/kaldi/egs/timit/s5$
```