

IE Project2

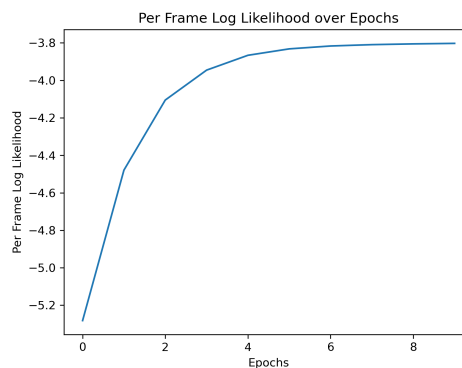
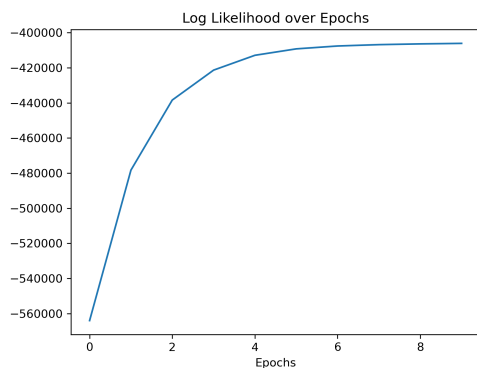
Yining Lu

March 2023

1 Primary System

1.1 Plot

```
echo: 0 log_likelihood = -564019.6755373772 per_frame_log_likelihood = -5.281824933627169
echo: 1 log_likelihood = -478297.31876903615 per_frame_log_likelihood = -4.47906839695684
echo: 2 log_likelihood = -438360.52005632914 per_frame_log_likelihood = -4.105075807054635
echo: 3 log_likelihood = -421310.9859881588 per_frame_log_likelihood = -3.9454135504814234
echo: 4 log_likelihood = -412858.8522180201 per_frame_log_likelihood = -3.866262604467108
echo: 5 log_likelihood = -409199.2630647716 per_frame_log_likelihood = -3.8319919751348186
echo: 6 log_likelihood = -407560.8167783121 per_frame_log_likelihood = -3.8166485627973223
echo: 7 log_likelihood = -406791.570444423 per_frame_log_likelihood = -3.809444870013794
echo: 8 log_likelihood = -406363.80770675157 per_frame_log_likelihood = -3.805439038317662
echo: 9 log_likelihood = -406073.64125771413 per_frame_log_likelihood = -3.802721742358141
```



The HMM was trained with 10 epochs and we can see that the likelihood had no significant increase after the 5th (i.e epoch = 4) epoch.

1.2 Test

The most likely words and their confidence:

After 10 Epochs

```
['enjoy', 'enjoy', 'extra', 'system', 'enjoy', 'along', 'enjoy', 'enjoy', 'enjoy', 'dinner',
'oil', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'never', 'enjoy', 'enjoy', 'another',
'enjoy', 'system', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'children', 'enjoy', 'brother',
'another', 'system', 'children', 'enjoy', 'enjoy', 'always', 'enjoy', 'enjoy', 'greasy',
```



```

1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.99994329 1. 0.8458762
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.9500686 1. 1. 0.99997253 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.99992477 1. 1. 0.99985874 1. 1.
1. 1. 1. 1. 0.99999979 1. 1. 1. 1. 1. 1. 0.88198536 1. 1. 1. 1. 1. 1.
0.64591577 1. 1. 1. 1. 1. 0.99956827 1. 1. 0.95085968 1. 1. 0.99947252 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 0.99894577 1. 1. 1. 1. 0.99999999 1. 0.99999981
1. 1. 1. 1. 1. 1. 1. 0.99998924 1. 1. 1. 1. 1. 1. 0.99999987 0.99999999
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. ]

```

1.3 Source Code

Here is the command I used to run my model: `python project2.py`. For more implementation details, please refer to `README.md`.

2 Held-Out Accuracy Test

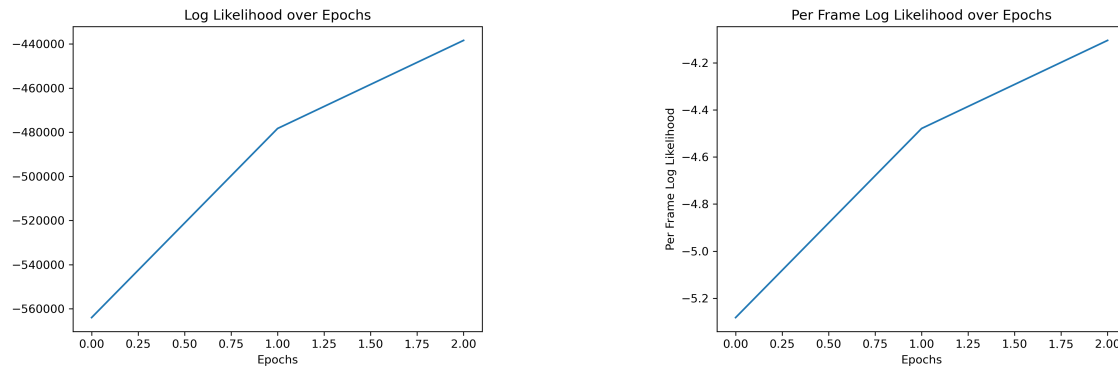
$N^* = 3$ training epochs are needed.

```

Kept set size: 640
Held-out set size: 158
echo: 0 log_likelihood = -452417.0465016624 per_frame_log_likelihood = -5.284073003675147
Hold-out accuracy: 0.620253164556962
echo: 1 log_likelihood = -383651.2833434081 per_frame_log_likelihood = -4.480912920536424
Hold-out accuracy: 0.6962025316455697
echo: 2 log_likelihood = -351133.3934218824 per_frame_log_likelihood = -4.101115329796919
Hold-out accuracy: 0.7215189873417721
echo: 3 log_likelihood = -336822.21892920183 per_frame_log_likelihood = -3.9339658128359574
Hold-out accuracy: 0.689873417721519

```

2.1 Plot



2.2 Test

After 3 Epochs

```

['eating', 'become', 'also', 'around', 'problem', 'enough', 'along', 'over', 'very',
'always', 'oily', 'along', 'always', 'money', 'always', 'only', 'never', 'nothing', 'also',

```

'after', 'around', 'another', 'problem', 'over', 'morning', 'carry', 'around', 'eating',
 'brother', 'another', 'about', 'water', 'often', 'nothing', 'enough', 'money', 'sometimes',
 'greasy', 'every', 'several', 'morning', 'oily', 'dinner', 'forces', 'study', 'without',
 'over', 'before', 'people', 'carry', 'only', 'even', 'only', 'over', 'without', 'water',
 'carry', 'oily', 'over', 'oily', 'many', 'before', 'before', 'shellfish', 'forces', 'money',
 'people', 'greasy', 'before', 'every', 'very', 'people', 'government', 'problem', 'government',
 'also', 'every', 'away', 'also', 'another', 'very', 'dinner', 'greasy', 'water', 'carry',
 'water', 'morning', 'about', 'around', 'people', 'children', 'become', 'also', 'never',
 'also', 'money', 'only', 'over', 'many', 'problem', 'brother', 'about', 'children', 'many',
 'money', 'other', 'away', 'oily', 'money', 'oily', 'oily', 'without', 'brother', 'money',
 'problem', 'often', 'about', 'dinner', 'many', 'water', 'only', 'study', 'shellfish',
 'always', 'never', 'about', 'along', 'order', 'never', 'allow', 'very', 'carry', 'away',
 'many', 'sometimes', 'problem', 'become', 'about', 'without', 'after', 'around', 'people',
 'before', 'even', 'always', 'oily', 'carry', 'away', 'greasy', 'eating', 'into', 'very',
 'several', 'shellfish', 'over', 'many', 'become', 'other', 'about', 'without', 'water',
 'never', 'order', 'only', 'carry', 'open', 'about', 'oily', 'enough', 'very', 'brother',
 'without', 'water', 'often', 'sometimes', 'very', 'forces', 'only', 'money', 'along',
 'nothing', 'along', 'many', 'morning', 'many', 'often', 'table', 'very', 'oily', 'other',
 'after', 'carry', 'open', 'shellfish', 'another', 'order', 'become', 'along', 'never',
 'never', 'greasy', 'oily', 'money', 'greasy', 'often', 'without', 'government', 'many',
 'after', 'water', 'over', 'enough', 'only', 'only', 'greasy', 'eating', 'after', 'often',
 'greasy', 'system', 'many', 'water', 'before', 'very', 'very', 'many', 'sometimes', 'into',
 'sometimes', 'shellfish', 'system', 'always', 'into', 'never', 'several', 'many', 'after',
 'after', 'system', 'money', 'another', 'oily', 'around', 'order', 'many', 'often', 'greasy',
 'away', 'people', 'eating', 'brother', 'sometimes', 'greasy', 'enough', 'open', 'people',
 'never', 'order', 'often', 'enjoy', 'often', 'before', 'oily', 'another', 'children',
 'about', 'very', 'always', 'greasy', 'greasy', 'government', 'very', 'never', 'also',
 'problem', 'often', 'problem', 'about', 'many', 'money', 'study', 'also', 'people', 'study',
 'after', 'about', 'after', 'along', 'very', 'children', 'after', 'dinner', 'often', 'enough',
 'greasy', 'along', 'order', 'every', 'children', 'very', 'oily', 'enjoy', 'always', 'before',
 'morning', 'open', 'only', 'water', 'table', 'over', 'about', 'another', 'many', 'oily',
 'money', 'another', 'people', 'along', 'enough', 'oily', 'carry', 'morning', 'always',
 'extra', 'government', 'extra', 'without', 'water', 'carry', 'nothing', 'before', 'every',
 'only', 'along', 'open', 'before', 'very', 'without', 'often', 'always', 'forces', 'after',
 'very', 'along', 'over', 'table', 'many', 'children', 'allow', 'about', 'every', 'forces',
 'after', 'only', 'even', 'away', 'always', 'carry', 'system', 'children', 'over', 'carry',
 'money', 'oily', 'even', 'oily', 'often', 'many', 'also', 'allow', 'children', 'also',
 'along', 'many', 'very', 'without', 'many', 'people', 'eating', 'into', 'never', 'without',
 'become', 'oily', 'system', 'enjoy', 'other', 'extra', 'before', 'enjoy', 'shellfish',
 'after', 'greasy'] [0.52893098 1. 1. 1. 1. 0.99996517 0.93654027 1. 0.89173843 0.68488075
 0.99808334 0.99161935 0.99999909 0.99955618 0.99477393 0.51666283 1. 1. 0.99999988
 0.99920765 0.99997264 1. 0.99997964 1. 1. 1. 0.99976392 0.96477255 0.99999961 0.99995984
 1. 1. 0.99992333 1. 0.99963662 0.99995689 1. 0.99999464 0.87513152 1. 1. 0.99999832
 0.99973419 1. 0.99999981 0.81892237 0.83157517 0.99999991 1. 1. 0.99999848 0.9999907
 0.98971727 0.97727093 1. 1. 1. 1. 1. 0.84397484 0.99999996 1. 1. 1. 1. 0.99997866
 0.9999779 1. 1. 0.93680011 0.99078452 1. 1. 1. 1. 1. 0.99996077 0.99994392 1.
 0.99999996 0.6713422 1. 0.99996882 1. 1. 0.98506921 1. 0.99999999 1. 0.99999959
 1. 1. 1. 1. 1. 0.9999989 0.9954698 0.99035321 0.98955769 1. 1. 1. 0.99972206
 0.99999718 0.94301566 0.99999047 0.99747094 0.73940252 0.7075102 0.74388926 0.67910431
 0.99999829 1. 0.99993909 1. 0.99996933 1. 0.94740093 0.99999997 1. 0.95094236 1.
 0.99800795 0.99984147 1. 1. 0.9999997 0.99006437 1. 0.98122269 0.93518942 1. 0.99992869

```

0.99870342 1. 0.9888551 1. 1. 0.99999813 1. 1. 0.99981414 0.9997987 0.870062 1.
0.92367573 1. 0.99982179 1. 1. 0.99999977 0.91869886 1. 1. 1. 0.99999992 0.73122613
0.99967372 1. 1. 1. 1. 0.99999979 0.87424478 1. 0.99999446 1. 0.64099084 0.55308419
0.97547291 1. 1. 1. 0.99994494 0.99999998 0.99072409 0.785266 0.92888831 0.95650457
0.99960392 1. 1. 0.99983548 0.99999089 0.99855927 1. 1. 0.96272794 0.95334359 0.99999094
0.9989399 1. 0.9999762 1. 1. 0.99999982 1. 0.99924929 0.99999795 1. 0.99998451
0.71492224 0.99999115 1. 0.9831312 0.99998677 1. 0.98600993 0.78274625 1. 0.99646409
0.99977373 0.99106414 0.91869138 1. 1. 0.99999864 0.99999804 1. 0.99997764 0.72227837
1. 0.97476239 0.98643619 0.97171017 0.99985995 1. 0.9985957 0.99987348 1. 0.99977273
0.9999977 1. 1. 0.99997433 0.9928511 0.99994622 1. 1. 0.94252058 1. 0.54005027 1.
0.70590026 0.99987896 0.99967017 0.95631877 0.99999427 0.99979747 1. 1. 1. 1.
0.99992246 1. 1. 0.93904089 0.99104986 1. 1. 1. 0.99982295 1. 1. 0.99999988 0.74619685
0.98193672 0.99677024 1. 1. 0.97319044 1. 1. 0.99358691 0.99970519 1. 1. 0.99999984
0.42857294 0.99999867 0.62214913 0.59258368 1. 0.99999993 0.99998688 0.99999294 0.99989524
0.86789912 1. 0.99999994 0.99976381 0.9994695 1. 1. 0.92182795 1. 0.66947339 1. 0.98326755
1. 1. 0.99999952 1. 1. 0.99999999 0.8932611 0.99999945 1. 0.99999148 0.99672032
0.99744816 1. 0.99999985 0.99679787 0.99945685 0.99999978 0.80454766 1. 0.9903903 1.
0.99999784 0.99999816 0.99999998 1. 0.92129586 0.99999984 0.87601358 1. 1. 0.99999976
0.50370078 0.9627593 0.99998099 0.99999984 1. 0.52793996 1. 0.92977692 0.99999999 1.
0.99999994 0.95971048 0.36446611 0.9954953 1. 0.90598022 1. 0.99997868 0.86242547 0.98180647
1. 1. 0.61633329 0.97128343 0.99931297 0.99999475 0.99714604 1. 1. 0.96692343 1.
0.99992432 0.97685542 0.99998754 0.97373638 0.99999989 0.99940948 1. 0.62051332 1. 1.
0.53360926 0.99678404 0.98688378 1. 0.96760648 1. 1. 1. 1. 0.4547076 1. 0.81984032
1. 1. 0.77989553 0.99999998 0.99999727 1. 1. 0.99999343 1. ]

```

2.3 Source Code

The code file is `project2_accuracy.py`. To find the optimal number of iterations, run `python project2_accuracy.py --find_N --num_epochs 10`. To train the HMM with $N^* = 3$ and to reproduce the above result, run `python project2_accuracy.py --num_epochs 3`. More implementation details are in `README.md`

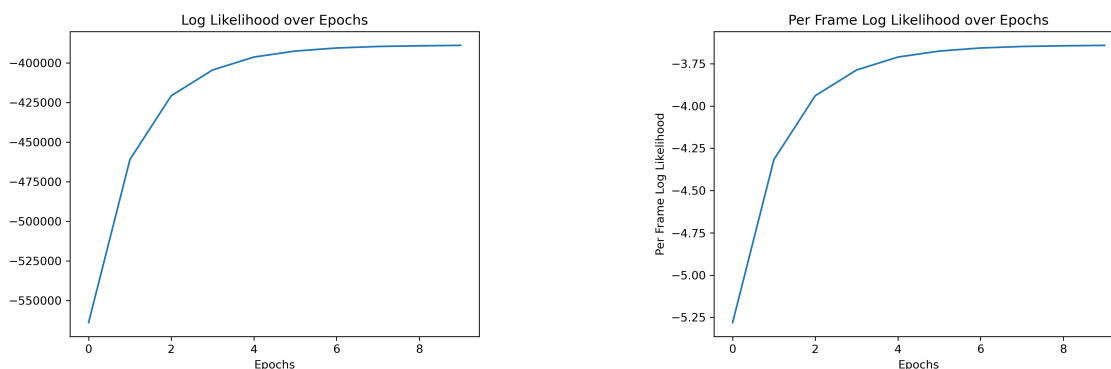
3 Phonemic Baseforms

3.1 Plot

```

echo: 0 log_likelihood = -563892.3192077931 per_frame_log_likelihood = -5.280632291125094
echo: 1 log_likelihood = -460941.87943497376 per_frame_log_likelihood = -4.316541456524547
echo: 2 log_likelihood = -420654.14642795705 per_frame_log_likelihood = -3.939262503422363
echo: 3 log_likelihood = -404408.13045916223 per_frame_log_likelihood = -3.7871248813893548
echo: 4 log_likelihood = -396260.368806244 per_frame_log_likelihood = -3.710824261893
echo: 5 log_likelihood = -392470.0132320722 per_frame_log_likelihood = -3.6753290558793106
echo: 6 log_likelihood = -390527.3283023449 per_frame_log_likelihood = -3.657136566955517
echo: 7 log_likelihood = -389575.9460144208 per_frame_log_likelihood = -3.648227241788836
echo: 8 log_likelihood = -389146.1087313496 per_frame_log_likelihood = -3.644201982781754
echo: 9 log_likelihood = -388883.0132553676 per_frame_log_likelihood = -3.64173819595793

```



The HMM trained with phonemic baseforms demonstrates a slightly higher likelihood after 10 epochs when compared to the HMM in the primary system. But it also stopped increasing the likelihood after about 5 epochs.

3.2 Test

After 10 Epochs ['enjoy', 'even', 'even', 'forces', 'forces', 'problem', 'even', 'government', 'shellfish', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'open', 'enjoy', 'even', 'problem', 'enjoy', 'every', 'enjoy', 'forces', 'enjoy', 'even', 'forces', 'forces', 'government', 'enjoy', 'problem', 'after', 'enjoy', 'even', 'forces', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'forces', 'enjoy', 'even', 'even', 'even', 'even', 'even', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'shellfish', 'shellfish', 'enjoy', 'forces', 'enjoy', 'enjoy', 'forces', 'even', 'enjoy', 'enjoy', 'forces', 'even', 'enjoy', 'forces', 'enjoy', 'forces', 'enjoy', 'enjoy', 'forces', 'forces', 'enjoy', 'forces', 'enjoy', 'enjoy', 'always', 'enjoy', 'enjoy', 'enjoy', 'always', 'enjoy', 'enjoy', 'enjoy', 'forces', 'even', 'forces', 'forces', 'enjoy', 'forces', 'even', 'enjoy', 'forces', 'even', 'shellfish', 'enjoy', 'enjoy', 'enjoy', 'forces', 'forces', 'enjoy', 'forces', 'forces', 'enjoy', 'government', 'enjoy', 'forces', 'enjoy', 'enjoy', 'problem', 'forces', 'shellfish', 'enjoy', 'forces', 'enjoy', 'enjoy', 'problem', 'enjoy', 'shellfish', 'enjoy', 'enjoy', 'enjoy', 'even', 'forces', 'enjoy', 'enjoy', 'shellfish', 'shellfish', 'enjoy', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy', 'government', 'even', 'enjoy', 'even', 'enjoy', 'enjoy', 'enjoy', 'shellfish', 'forces', 'enjoy', 'even', 'enjoy', 'after', 'enjoy', 'forces', 'enjoy', 'even', 'enjoy', 'even', 'even', 'enjoy', 'enjoy', 'enjoy', 'forces', 'enjoy', 'even', 'problem', 'even', 'forces', 'enjoy', 'allow', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'forces', 'even', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'forces', 'enjoy', 'forces', 'enjoy', 'enjoy', 'even', 'forces', 'enjoy', 'enjoy', 'enjoy', 'shellfish', 'forces', 'shellfish', 'even', 'shellfish', 'enjoy', 'even', 'forces', 'problem', 'problem', 'even', 'enjoy', 'forces', 'even', 'enjoy', 'enjoy', 'enjoy', 'forces', 'every', 'even', 'shellfish', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'enjoy', 'forces', 'problem', 'even', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'problem', 'even', 'forces', 'enjoy', 'even', 'every', 'even', 'forces', 'enjoy', 'enjoy', 'shellfish', 'enjoy', 'enjoy', 'enjoy', 'forces', 'enjoy', 'even', 'enjoy', 'even', 'even', 'even', 'even', 'enjoy', 'shellfish', 'even', 'even', 'even', 'enjoy', 'forces', 'problem', 'shellfish', 'enjoy', 'forces', 'enjoy', 'problem', 'enjoy', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy', 'even', 'even', 'enjoy', 'forces', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'even', 'forces', 'even', 'enjoy', 'enjoy', 'sometimes', 'forces', 'shellfish', 'after', 'enjoy', 'even', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'after',

```

'even', 'enjoy', 'enjoy', 'enjoy', 'forces', 'open', 'forces', 'even', 'shellfish', 'enjoy',
'forces', 'even', 'problem', 'enjoy', 'enjoy', 'enjoy', 'forces', 'enjoy', 'enjoy', 'enjoy',
'even', 'even', 'enjoy', 'enjoy', 'enjoy', 'every', 'forces', 'even', 'enjoy', 'enjoy',
'enjoy', 'forces', 'shellfish', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even',
'enjoy', 'enjoy', 'enjoy', 'shellfish', 'enjoy', 'enjoy', 'even', 'enjoy', 'enjoy', 'enjoy',
'even', 'shellfish', 'even', 'enjoy', 'forces', 'even', 'enjoy', 'enjoy', 'forces', 'enjoy',
'enjoy', 'enjoy', 'enjoy', 'enjoy', 'enjoy', 'even', 'enjoy', 'enjoy', 'enjoy', 'forces',
'even', 'problem', 'problem', 'enjoy', 'open', 'enjoy', 'forces', 'forces', 'enjoy',
'forces', 'shellfish', 'even', 'even', 'forces', 'forces', 'enjoy', 'enjoy', 'enjoy',
'after', 'forces', 'problem', 'enjoy'] [1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]

```

3.3 Source Code

The code file is `project2_pronunciation.py` and the one extra dataset file I used is `./data_phonemes/cmudict-0.7b.txt` downloaded from the official website. To run the train and test module, use the command `python project2_pronunciation.py`. For more code details, please refer to `README.md`