Solution:
All below model is trained on ‘wnut17train.conll’, and tested on ‘emerging.test.conll’

a) CRF trained with default setting obtained avg f1_score on test data of 0.916.

```
                 precision    recall  f1-score   support

               O      0.945     0.996     0.970     21934
   B-corporation      0.000     0.000     0.000       117
   I-corporation      0.000     0.000     0.000        24
 B-creative-work      0.500     0.013     0.025       231
 I-creative-work      0.786     0.046     0.088       237
         B-group      0.000     0.000     0.000        97
         I-group      0.143     0.026     0.043        39
      B-location      0.235     0.189     0.209       122
      I-location      0.179     0.128     0.149        39
        B-person      0.516     0.093     0.158       355
        I-person      0.515     0.168     0.254       101
       B-product      0.667     0.036     0.068        56
       I-product      0.250     0.024     0.043        42

        accuracy                          0.938     23394
       macro avg      0.364     0.132     0.154     23394
    weighted avg      0.913     0.938     0.916     23394
```

b) Softmax classifier trained with default setting, obtained avg f1_score on test data of 0.907, where CRF performed better with a higher f1_score. In part c, I will proceed to fine tune hyperparameters of the classifier.

```
                 precision    recall  f1-score   support

               O      0.938     1.000     0.968     21934
   B-corporation      0.000     0.000     0.000       117
   I-corporation      0.000     0.000     0.000        24
 B-creative-work      0.000     0.000     0.000       231
 I-creative-work      0.000     0.000     0.000       237
         B-group      0.000     0.000     0.000        97
         I-group      0.000     0.000     0.000        39
      B-location      0.000     0.000     0.000       122
      I-location      0.000     0.000     0.000        39
        B-person      0.000     0.000     0.000       355
        I-person      0.000     0.000     0.000       101
       B-product      0.000     0.000     0.000        56
       I-product      0.000     0.000     0.000        42

        accuracy                          0.938     23394
       macro avg      0.072     0.077     0.074     23394
    weighted avg      0.879     0.938     0.907     23394
```

c) Softmax classifier hyperparameter tuning: test avg f1_score = 0.952

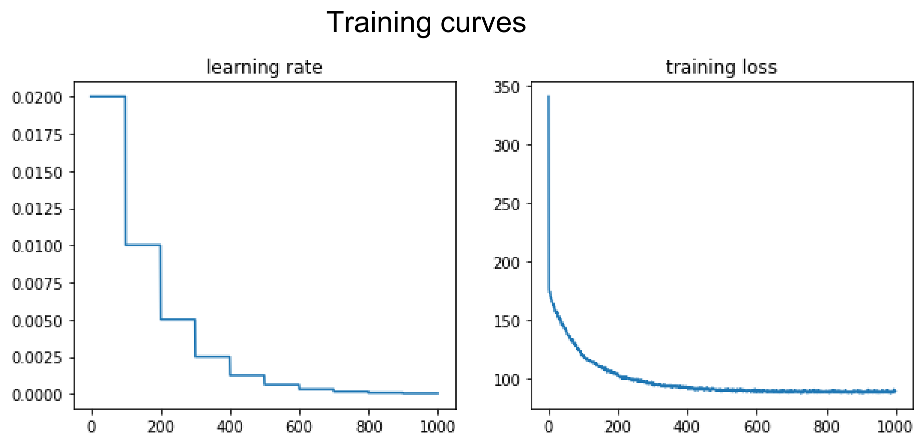
I tried below values for each hyperparameter, and the avg f1 score on the dev dataset will be the criteria to choose the best parameter settings.

Table of hyperparameter tuning validation f1 score

Training config (parameters same as default other than indicated)	**Avg f1 score on dev dataset** **(**‘emerging.dev.conll’)
Default 'batch_size': 4, 'half_window': 2, 'embed_dim': 25, 'hidden_dim': 25, 'num_classes': 13, 'freeze_embeddings': False, 'lr': 0.02, 'epochs': 100 ‘Num_layers’: 1	0.8824671395802548
Batch_size 2	0.8786100406439366
Batch_size 8	0.8824671395802548
Embed_dim 50	0.8824181390344367
Embed_dim 100 **(2nd best)**	0.**8828044283450783**
Embed_dim 200	0.8824118855465907
Embed_dim 300	0.8822393988274964
Half window 3	0.8824036834877962
Half window 4	0.8824963458400458
Half window 5	0.8824354125841017
Hidden dim 50	0.8824671395802548
Hidden dim 100	0.8824631384713871
Number layer 2	0.8823476631930625
Number layer 3(**best**)	0.**8828561475496695**
Number layer 4	0.8823008878819346
Freeze_embeddings True	0.8823326493950854
Lr 0.002	0.8821357462243560
Lr 0.1	0.8814530901515277
Epochs 200	0.8829305762387519
Epochs 300(**improve with epochs**)	0.8831023763895041
Combine two parameters Embed dim 100, num layer 3	0.8820091892966707

From the above table, the most effective parameter is 3 hidden layers(default 1) plus more epochs. To get better results, I also used a learning rate scheduler to reduce the rate by half per 100 epochs. The model was trained for 1000 epochs(when loss converges) and got avg f1 score of 0.952 on test data.

The learning curve and classification report are attached below.



```
                  precision    recall  f1-score   support

               O      0.968     0.999     0.983     44296
   B-corporation      0.778     0.356     0.488       177
   I-corporation      0.000     0.000     0.000        36
 B-creative-work      0.471     0.075     0.129       107
 I-creative-work      0.743     0.310     0.437       168
         B-group      0.308     0.027     0.050       148
         I-group      0.444     0.037     0.068       108
      B-location      0.714     0.389     0.503       391
      I-location      0.741     0.244     0.367       164
        B-person      0.619     0.261     0.367       468
        I-person      0.727     0.213     0.330       225
       B-product      0.375     0.030     0.056       100
       I-product      0.833     0.062     0.115        81

        accuracy                          0.963     46469
       macro avg      0.594     0.231     0.299     46469
    weighted avg      0.952     0.963     0.952     46469
```