Project 5

Kyle McClintick
ECE 579

May 7, 2020

The following sections document the training of various models on the 6000-feature PerfWeb classification of 40 incognito Tor websites by printing grid-search validation results. Conclusions of results are presented at the end.

1 SVM

```
( 1 / 12 ) validation CE acc for h ['linear' '0.01'] :
0.2975
( 2 / 12 ) validation CE acc for h ['linear' '0.1'] :
0.43
( 3 / 12 ) validation CE acc for h ['linear' '1.0'] :
0.575
( 4 / 12 ) validation CE acc for h ['poly' '0.01'] : 0.185
( 5 / 12 ) validation CE acc for h ['poly' '0.1'] : 0.185
( 6 / 12 ) validation CE acc for h ['poly' '1.0'] : 0.185
( 7 / 12 ) validation CE acc for h ['rbf' '0.01'] : 0.275
( 8 / 12 ) validation CE acc for h ['rbf' '0.1'] : 0.275
( 9 / 12 ) validation CE acc for h ['rbf' '1.0'] : 0.275
( 10 / 12 ) validation CE acc for h ['sigmoid' '0.01'] :
0.27
( 11 / 12 ) validation CE acc for h ['sigmoid' '0.1'] :
0.27
( 12 / 12 ) validation CE acc for h ['sigmoid' '1.0'] :
0.27
```

2 KNN

```
( 1 / 8 ) validation CE acc for h ['1' 'uniform'] : 0.69
( 2 / 8 ) validation CE acc for h ['1' 'distance'] : 0.69
( 3 / 8 ) validation CE acc for h ['5' 'uniform'] : 0.625
( 4 / 8 ) validation CE acc for h ['5' 'distance'] : 0.6525
( 5 / 8 ) validation CE acc for h ['20' 'uniform'] : 0.5225
( 6 / 8 ) validation CE acc for h ['20' 'distance'] :
0.5675
( 7 / 8 ) validation CE acc for h ['100' 'uniform'] :
0.365
( 8 / 8 ) validation CE acc for h ['100' 'distance'] :
0.4625
```

3 Decision Tree

```
( 1 / 4 ) validation CE acc for h ['best' 'entropy'] :
0.645
new best!
( 2 / 4 ) validation CE acc for h ['best' 'gini'] : 0.66
new best!
( 3 / 4 ) validation CE acc for h ['random' 'entropy'] :
0.6575
( 4 / 4 ) validation CE acc for h ['random' 'gini'] :
0.6425
```

4 CNN

```
( 1 / 16 ) validation acc for h [ 2. 2. 8. 10. 0.] :
0.6399999987334013
( 2 / 16 ) validation acc for h [ 2. 4. 8. 10. 0.] :
0.45250000022351744
( 3 / 16 ) validation acc for h [ 5. 2. 8. 10. 0.] :
0.5625000016763806
( 4 / 16 ) validation acc for h [ 5. 4. 8. 10. 0.] :
0.025
( 5 / 16 ) validation acc for h [ 2. 2. 64. 10. 0.] :
0.6750000005587935
```

```
( 6 / 16 ) validation acc for h  [ 2.  4. 64. 10.  0.] :
0.6700000004842878
( 7 / 16 ) validation acc for h  [ 5.  2. 64. 10.  0.] :
0.6699999982491136
( 8 / 16 ) validation acc for h  [ 5.  4. 64. 10.  0.] :
0.6375000011175871
( 9 / 16 ) validation acc for h  [ 2.  2.  8. 10.  0.5] :
0.32000000271946194
( 10 / 16 ) validation acc for h  [ 2.  4.  8. 10.
0.5] :  0.42250000294297935
( 11 / 16 ) validation acc for h  [ 5.  2.  8. 10.
0.5] :  0.030000000074505807
( 12 / 16 ) validation acc for h  [ 5.  4.  8. 10.
0.5] :  0.037500000186264516
( 13 / 16 ) validation acc for h  [ 2.  2. 64. 10.
0.5] :  0.5700000012293458
( 14 / 16 ) validation acc for h  [ 2.  4. 64. 10.
0.5] :  0.6424999982118607
( 15 / 16 ) validation acc for h  [ 5.  2. 64. 10.
0.5] :  0.11750000044703483
( 16 / 16 ) validation acc for h  [ 5.  4. 64. 10.
0.5] :  0.0925000024959445
```

5 Bonus: LSTM RNN

```
( 1 / 2 ) validation CE acc for h  [ 0. 16. 10.  0.] :
0.05
( 2 / 2 ) validation CE acc for h  [ 0. 64. 10.  0.] :
0.025
```

6 Conclusions

While the CNN requires further hyper-parameter validation to achieve the 80 % accuracy reported in the publication, it performs among the best models tested. Others include the decision tree and KNN with $K = 1$. Given an accuracy floor of $\frac{1}{40} = 2.5\%$, the LSTM was very disappointing for its historical benefit in classifying temporal data sets. Further validation would be required, particularly for model depth/width.

Table 1: A summary of the test accuracy of the models studied.

Model	SVM	KNN	DT	CNN	LSTM
Test Acc	57.5	69.0	66.0	67.5	5.0