Practical 2: Malware Classification

Ethan Alley, Grigory Khimulya, Walter Martin
alley@college.harvard.edu, khimulya@college.harvard.edu, wmartin@college.harvard.edu

March 9, 2017

1 Technical Approach

We explored the problem of malware classification from several different angles:

- Running through a variety of models quickly to test which show the most immediate promise
- Feature engineering
- More focused tuning of hyperparameters for neural nets

How did you tackle the problem? Credit will be given for:

- Diving deeply into a method (rather than just trying off-the-shelf tools with default settings). This can mean providing mathematical descriptions or pseudo-code.
- Making tuning and configuration decisions using thoughtful experimentation. This can mean carefully describing features added or hyperparameters tuned.
- Exploring several methods. This can contrasting two approaches or perhaps going beyond those we discussed in class.

Thoughtfully iterating on approaches is key. If you used existing packages or referred to papers or blogs for ideas, you should cite these in your report.

2 Results

This section should report on the following questions:

- Did you create and submit a set of predictions?
- Did your methods give reasonable performance?

You must have *at least one plot or table* that details the performances of different methods tried. Credit will be given for quantitatively reporting (with clearly labeled and captioned figures and/or tables) on the performance of the methods you tried compared to your baselines.

Mention Features	
Feature	Value Set
Mention Head	$\mathcal{V}$
Mention First Word	$\mathcal{V}$
Mention Last Word	$\mathcal{V}$
Word Preceding Mention	$\mathcal{V}$
Word Following Mention	$\mathcal{V}$
# Words in Mention	$\{1, 2, \ldots\}$
Mention Type	$\mathcal{T}$

Table 1: Feature lists are a good way of illustrating problem specific tuning.

Model	Acc.
BASELINE 1	0.45
BASELINE 2	2.59
MODEL 1	10.59
MODEL 2	13.42
MODEL 3	7.49

Table 2: Result tables can compactly illustrate absolute performance, but a plot may be more effective at illustrating a trend.

3 Discussion

End your report by discussing the thought process behind your analysis. This section does not need to be as technical as the others but should summarize why you took the approach that your did. Credit will be given for:

- Explaining the your reasoning for why you seqentially chose to try the approaches you did (i.e. what was it about your initial approach that made you try the next change?).
- Explaining the results. Did the adaptations you tried improve the results? Why or why not? Did you do additional tests to determine if your reasoning was correct?