PLS [3] are a no-regret generalization of HSM to multi-label problems.

- Extended code $z = (1, ..., h)$

- Factorization of the marginal probability:
  \[
  \psi(y) = P(z) = \prod_i P(z_i | x) = \prod_i \left( \frac{1}{\sum_i P(z_i | x)} \right) P(z_i | x)
  \]

- Different normalization than in HSM:
  \[
  \psi(y) \geq 1
  \]

- PLSs to a multi-class distribution boil down to HSM

Regret bounds

- Bound for the absolute difference between the true and the estimated marginal probability for label $l$:
  \[
  \|\psi(y) - y_l\|_1 = \sum_i P(z_i | x) \geq 1
  \]

- Bound for the regret with respect to precision@k:
  \[
  \|\psi(y) - y_l\|_1 \geq \sum_i P(z_i | x) \leq 1
  \]

Implementation (extremeText)

- Based on fastText
- Tree structure: random, Huffman tree or build via top-down hierarchical balanced clustering
- Linear models in the nodes
- Online training with features embedding (hidden, dense representation)
- $L_2$ regularization for all parameters of the model (for embedding and internal node classifiers).
- Hidden representation obtained by weighted average of the feature vector of proportion to the height of the tree for each feature.
- Depth-first search prediction for fast online prediction.

Source code: https://github.com/mwydmuch/extremeText

Experimental results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>train</th>
<th>test</th>
<th>Amazon-670K</th>
<th>Amazon-945K</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSM</td>
<td>1.5G</td>
<td>7032m</td>
<td>4.68ms</td>
<td>16.18ms</td>
</tr>
<tr>
<td>Huffman</td>
<td>7.5G</td>
<td>2.05</td>
<td>45.84</td>
<td>40.07</td>
</tr>
<tr>
<td>L2</td>
<td>3.2G</td>
<td>1.5G</td>
<td>58.70</td>
<td>61.53</td>
</tr>
<tr>
<td>fastText</td>
<td>1.9G</td>
<td>1.25</td>
<td>25.89</td>
<td>7.10</td>
</tr>
</tbody>
</table>

Ablation analysis for Amazon-670K

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>train</th>
<th>test</th>
<th>Amazon-670K</th>
<th>Amazon-945K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huffman</td>
<td>7.5G</td>
<td>2.05</td>
<td>45.84</td>
<td>40.07</td>
</tr>
<tr>
<td>L2</td>
<td>3.2G</td>
<td>1.5G</td>
<td>58.70</td>
<td>61.53</td>
</tr>
<tr>
<td>fastText</td>
<td>1.9G</td>
<td>1.25</td>
<td>25.89</td>
<td>7.10</td>
</tr>
</tbody>
</table>