

CS136

CP1 Pengcheng Xu

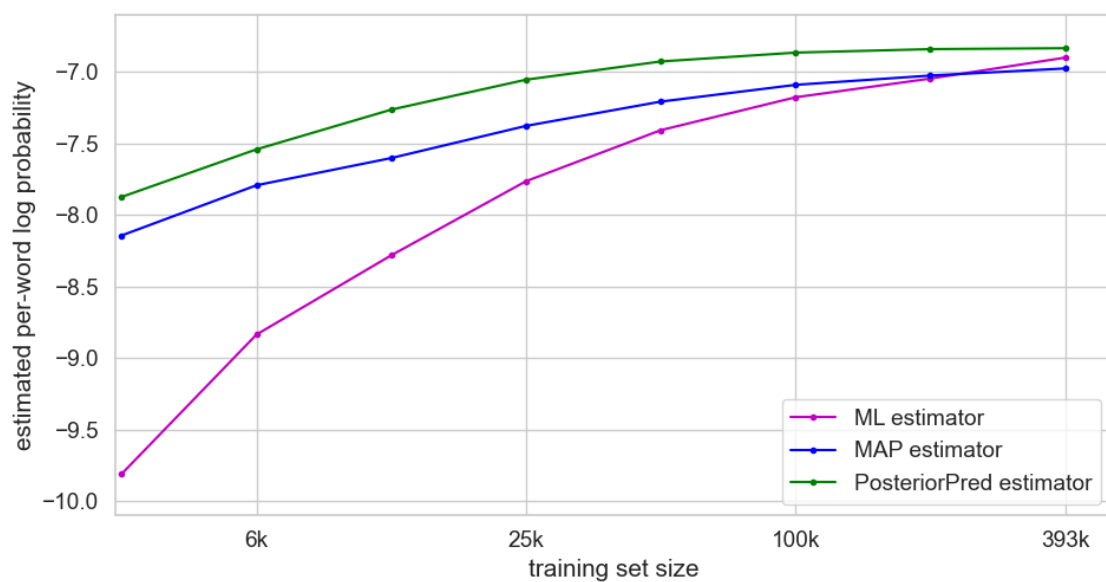
Collaboration Statement:

- Total hours spent: 4 hrs
- Consultd Resources:
 - Course's website
- Internet

Problem 1.

1a.

After implementing three estimators (i.e. ML, MAP, PosteriorPredictive) and setting the parameters (i.e. alpha, epsilon) given in the question description, we could get the figure below:



1b.

From the figure above, the ML estimator seems to eventually outperform other estimators when $N \rightarrow \infty$ based on their trends at the right edge of the figure.

However, from their formulas, we could tell that they eventually will converge to the same value.

(1) For ML estimator, $\frac{n_v}{N} \rightarrow \mu_v^*$, $N \rightarrow \infty$. /* we know ML estimator will converge to the true μ , which we denote as μ_v^* */

(2) For MAP estimator, $\frac{n_v + \alpha - 1}{N + V(\alpha - 1)} = \frac{\frac{n_v}{N} + \frac{(\alpha - 1)}{N}}{1 + \frac{V(\alpha - 1)}{N}} \rightarrow \frac{n_v}{N} = \mu_v^*$, $N \rightarrow \infty$. /* Since $\frac{\alpha - 1}{N}$ and $\frac{V(\alpha - 1)}{N} \rightarrow 0$, $N \rightarrow \infty$ */

(3) For PosterioriPredictive, the derivation is similar to (2)MAP estimator, except $\alpha - 1$ (i.e. in MAP) becomes α (i.e. in PosterioriPredictive), we could also have $\frac{n_v + \alpha}{N + V\alpha} \rightarrow \mu_v^*$.

1c.

If we use a uniform probability distribution over the vocabulary, then the calculation of the "score" could be reduced to $-\log(V)$ (i.e. $\frac{\sum_{i=1}^N \log p(X_i = x_i)}{N} = \frac{\sum_{i=1}^N \log \frac{1}{V}}{N} = \log \frac{1}{V}$).

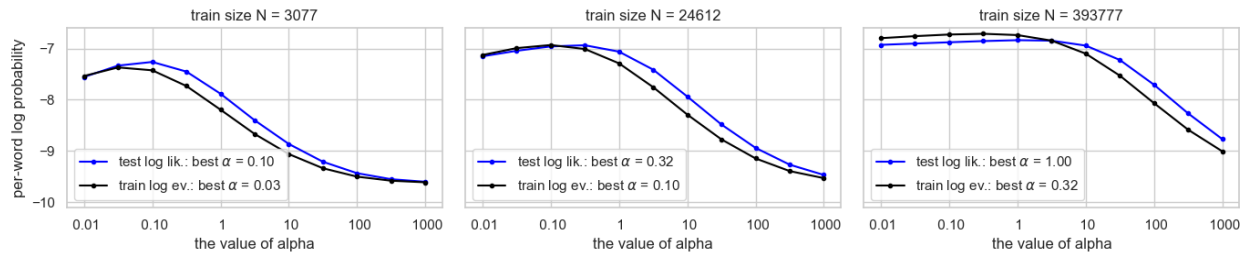
From the running `run_compare_estimators.py`, we know that V (i.e. vocab size) = 15205, so the score of uniform distribution = $-\log V = -9.629$.

From the figure we know that ML estimator actually does worse (i.e. close to -10.0) than the baseline when training data set is small.

Problem 2.

2a.

After implementing two parts (i.e. `calc_per_word_evidence` method and `__main__` section) in `run_select_alpha.py`, we have the figure below.



2b.

One advantage of "evidence-on-train" approach:

- Easy to apply. Don't need extra work to decide which part should be the validation set as in cross-validation approach.

One advantage of "conditional-likelihood-on-validation" approach:

- Avoid the overfitting to some degree. Since the idea is similar to cross-validation, which could avoid our model's performance depending on a specific selection of training and testing set.