Huilin Tong, ID:1261574
COMP135 Project 1

- The error rate of my model is nearly $0.15 - 0.25$, in other words, accuracy is about $75\% - 85\%$. With the smoothing parameter becoming larger, error rate also increase as shown in the plot below. But just little changes in error.

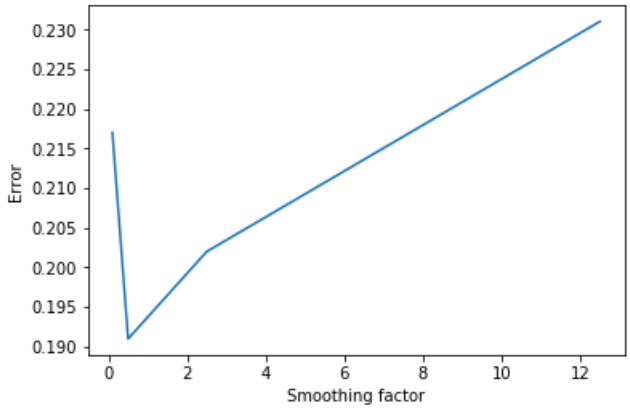

Fig.1 Amazon data

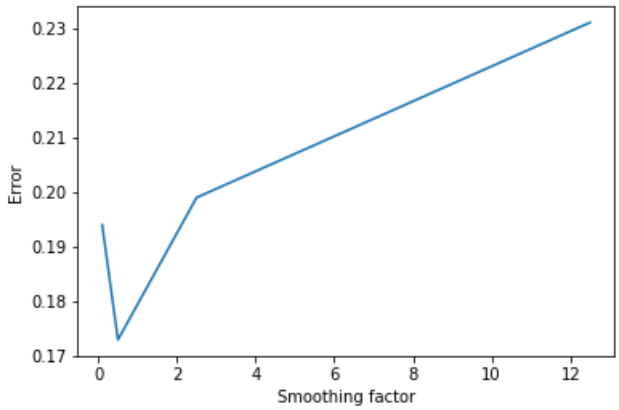

Fig.2 Yelp data

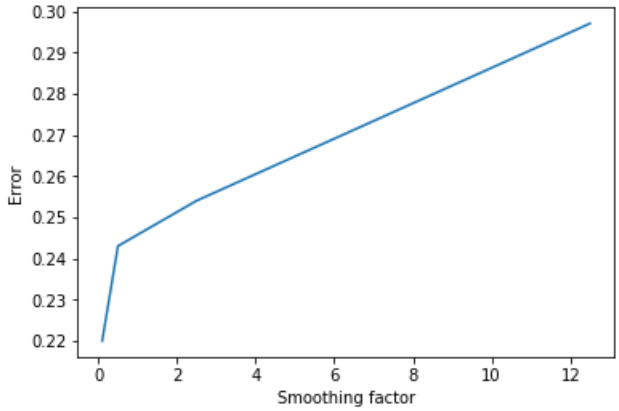

Fig.3 Imdb data

- When reducing the vocabulary size, error rate increases. This plot is only for smoothing parameter of $m = 0.1$.

  Large vocabulary size means the model is more accurate since you usually meet less words that not in the training set. So when you increase the vocabulary size, the error rate reduced

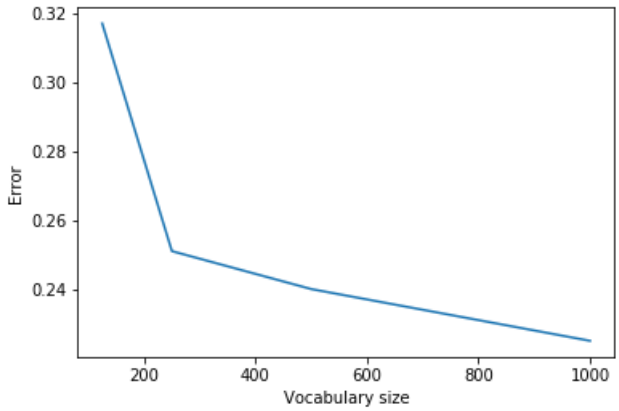

Fig.4 Amazon data

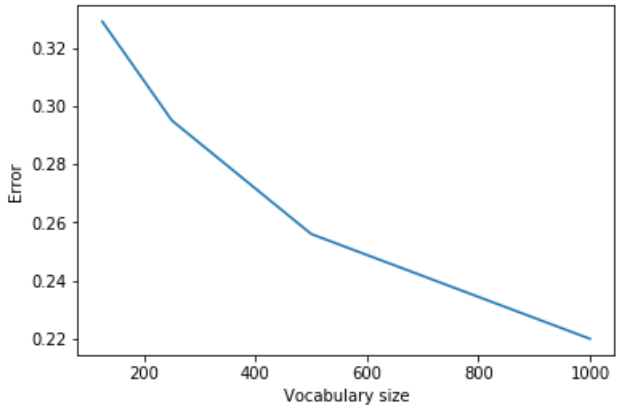

Fig.5 Yelp data

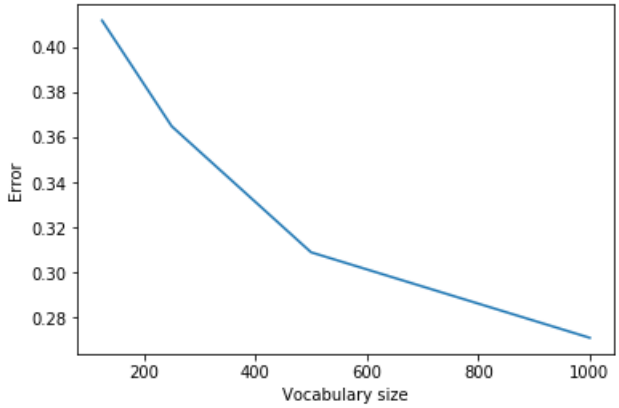

Fig.6 Imdb data

- Optional

Blue line represents for the validation error rate against the ration of training set. And red line stands for the real test error rate. This plot is only for the amazon data with smoothing parameter of $m = 0.1$. Code can show more. When training data is at the ration of 0.95, gives the best test error as shown in the plot.