CS771A
Introduction to Machine Learning
Group : ML_Alliance
Assignment 2

Jetha Ram	Sudhanshu Kumar	Deepali Singh
210473	211065	210309
Aakanksha Kumari	Pooja Kumari	Abhinav Singhal
210006	210726	210031

Problem 1.1

Give detailed calculations explaining the various design decisions you took to develop your ML algorithm. For instance, if you did use a decision tree algorithm, this includes the criterion to choose the splitting criterion at each internal node (e.g. if a certain bigram is present in the bigram list or not), criterion to decide when to stop expanding the decision tree and make the node a leaf, pruning strategies, hyperparameters etc.

Solution 1.1

Let us take a string **"beside"**

1) The set of bigrams are calculated by word "beside" - be, es, si, id, de 2) Duplicates should be removed

- In this example, two duplicates are found.

3) Lexicographic order
be, es, si, id, de - be, de, es, id, si
4) First 's' bigram is taken, so

- In this example, only "s" bigram is taken as it is: be, de, es, id, si

Now we have to predict words with the help of these 5 bigrams.
Case I: Let no duplicate bigram be found

- We make words (a part of words uses it)

- Start with "be" from the second phone should start with "e".

- be_

    - cannot take 'de' as it starts with 'd'

- – 'es' can be used

- bes_

  - – cannot take 'de' as it should start with 's'
  - – 'si' can be used

- besi_

  - – 'id'

- beside

Case 2

If its length is greater than 5, then it will be a part of the word.

Case 3

Duplicate is present.

Solution

At the time of searching, we go word by word and if the next letter does not match, we continue till the end.

Example: Searching from Dictionary

- according

- ccording

- cording

- ording

- rding

- ding

- ing

- ng

- g

We search like this for words, arrange the words by percentage of matching, and return the top 5 words.