Q1: (PAPERWORK) User enters text:

*in dairy life*

You should calculate probability if we put instead of dairy following words:

dairy, daisy, daily, fairy, fail

Use count\_1w.txt and count\_2w.txt

The probability of writing the word correctly 0.95 and incorrectly 0.05.

Assume that deletion of any letter is 0.005, insertion is also 0.005. substitution is 0.005.

Q2: (PAPERWORK) Find the probability of message S1 being spam and not spam using Laplace smoothing:

|  |  |
| --- | --- |
| **Training set** | **Input message** |
| **Spam messages:**  M1: *Buy bicycles for free*  M2: *Bicycles and motorbikes for free*  M3: *Motorbikes rides easy and free*  **Normal messages:**  M4: *Let's go ride bicycles*  M5: *Last week I bought motorbikes and they are cool*  M6: *Some messages about bicycles and motorbikes, that are free, are spam messages* | *Cool bicycles and motorbikes* |

Use Naive Bayes to find the probability of message Message being spam.

Q3: (IMPLEMENT CODE) Same task as Q1 but implemented on a computer for all words from count\_1w. (OPTIONAL) Use data from <http://norvig.com/ngrams/count_1edit.txt> to define error rates.