Aayush Shah

19BCE245

17 March 2021

Practical 7
Word probability

• **Definition :** Scan an integer n from the user. Scan n sentences (no special character, punctuation, all lower case letters). Each sentence's sentiment is either positive or negative. Scan 2 different words y and z from the user. y and z should be words from the set of distinct words constructed from n sentences. Write a program to estimate (i) the probability of y being present in the positive sentences (ii) the joint probability of y and z being present in positive sentences assuming that the presence of y and z in any sentence are independent events. What would be this probability if these events are not independent?

• **Code :**

```python
1. import string
2.
3. def isValid(str):
4.    invalidChars = set(string.punctuation)
5.    if any(char in invalidChars for char in str) or
   any(char.isupper() for char in str): #any function will
   return True if atleast one of the characters is in
   invalidChars.
6.        return False
7.    else:
8.        return True
9.
10.sentences = []
11.words = []
12.positive_words = []
13.negative_words = []
14.sentiments = []
15.
16.n = int(input("Enter number of sentences : "))
17.
```

```python
18. for index in range(n):
19.   sentences.append(str(input("\tEnter sentence {} : ".format(index+1))))
20.   while not isValid(sentences[index]):
21.       sentences[index] = str(input(f"\t\tEnter proper sentence {index+1} : "))
22.   words.extend(sentences[index].split())
23.
24.   sentiments.append(int(input("\tEnter sentiment of the sentence [1 for positive; 2 for negative] : ")))
25.   while sentiments[index] not in [1,2]:
26.       sentiments[index] = int(input(f"\t\tEnter proper sentiment : "))
27.
28.   if sentiments[index] == 1:
29.       positive_words.extend(sentences[index].split())
30.   else:
31.       negative_words.extend(sentences[index].split())
32.
33. distinct_words = set(words)
34. #print(distinct_words)
35.
36. y = str(input("Enter 1st word (y) : "))
37. while y not in distinct_words:
38.   y = str(input("\tEnter proper 1st word (y) : "))
39.
40. z = str(input("Enter 2nd word (z) : "))
41. while z not in distinct_words:
42.   z = str(input("\tEnter proper 2nd word (z) : "))
43.
44. p_of_y_present = positive_words.count(y)/ len(positive_words)
45. p_of_yz_present_independent = (positive_words.count(y)/ len(positive_words)) * (positive_words.count(z)/ len(positive_words))
46.
47. ans = 0
48. for index in range(n):
49.   if sentiments[index]==1 and (y in sentences[index]) and (z in sentences[index]):
50.       ans+=sentences[index].count(z)
51.
```

```python
52. p_of_yz_present_dependent = (ans*positive_words.count(y))/
    (len(positive_words))

53. print("The probability of y being present in the positive
    sentences is",p_of_y_present)
54. print("The probability of y and z being present
    [INDEPENDENT] in the positive sentences
    is",p_of_yz_present_independent)
55. print("The probability of y and z being present
    [  DEPENDENT] in the positive sentences
    is",p_of_yz_present_dependent)
```

- **Sample I/O :**

```
Enter number of sentences : 4
    Enter sentence 1 : Hello I'm Aayush
        Enter proper sentence 1 : hello i am aayush
    Enter sentiment of the sentence [1 for positive; 2 for negative] : 1
    Enter sentence 2 : feeling sad!
        Enter proper sentence 2 : feeling sad
    Enter sentiment of the sentence [1 for positive; 2 for negative] : 2
    Enter sentence 3 : aayush hello bro
    Enter sentiment of the sentence [1 for positive; 2 for negative] : 1
    Enter sentence 4 : aayush thats not good brother

    Enter sentiment of the sentence [1 for positive; 2 for negative] : 1
Enter 1st word (y) : aayush
Enter 2nd word (z) : hello
The probability of y being present in the positive sentences is 0.25
The probability of y and z being present [INDEPENDENT] in the positive sentences is 0.041666666666666664
The probability of y and z being present [  DEPENDENT] in the positive sentences is 0.5
```

```
✓ Run Succeeded    Time 75 ms    Peak Memory 7.5M                                        ƒ isValid ⌄   Tabs: 4 ⌄   Line 61, Column 119
```