191014012 - AI Classwork - Naïve Bayesian Classifier

December 2, 2021

Author Md. Azizul Hakim (191014012) azizul.hakim.cse@ulab.edu.bd

0.0.1 Words in My Dictionary

[1]:
```
words_in_dict = ["this", "hello", "how", "offer", "buy",
                 "thanks", "shipping", "meet", "tomorrow", "lunch",
                 "spicy", "morning", "evening", "function", "party"]
```

0.0.2 Spam Set & Ham Set Dataset

[2]:
```
### Spam Mail & Ham Mail Probability
spam_prob = 0.3
ham_prob = 0.7
```

[3]:
```
spam_set = dict()
```

[4]:
```
spam_set[words_in_dict[0]] = 500
spam_set[words_in_dict[1]] = 100
spam_set[words_in_dict[2]] = 700
spam_set[words_in_dict[3]] = 900
spam_set[words_in_dict[4]] = 850
spam_set[words_in_dict[5]] = 600
spam_set[words_in_dict[6]] = 950
spam_set[words_in_dict[7]] = 20
spam_set[words_in_dict[8]] = 3
spam_set[words_in_dict[9]] = 0
spam_set[words_in_dict[10]] = 0
spam_set[words_in_dict[11]] = 0
spam_set[words_in_dict[12]] = 0
spam_set[words_in_dict[13]] = 0
spam_set[words_in_dict[14]] = 0
```

[5]:
```
ham_set = dict()
```

[6]:
```
ham_set[words_in_dict[0]] = 700
ham_set[words_in_dict[1]] = 40
ham_set[words_in_dict[2]] = 750
```

```
ham_set[words_in_dict[3]] = 10
ham_set[words_in_dict[4]] = 10
ham_set[words_in_dict[5]] = 600
ham_set[words_in_dict[6]] = 1
ham_set[words_in_dict[7]] = 300
ham_set[words_in_dict[8]] = 100
ham_set[words_in_dict[9]] = 70
ham_set[words_in_dict[10]] = 50
ham_set[words_in_dict[11]] = 40
ham_set[words_in_dict[12]] = 30
ham_set[words_in_dict[13]] = 70
ham_set[words_in_dict[14]] = 50
```

0.0.3 Number of Words in Spam Set and Spam Set Word Likelihood

[7]:
```
spam_set_words = 0
```

[8]:
```
for word in words_in_dict:
    spam_set_words += spam_set[word]
```

[9]:
```
spam_set_word_likelihood = dict()
```

[10]:
```
for word in words_in_dict:
    spam_set_word_likelihood[word] = spam_set[word] / spam_set_words
```

[11]:
```
spam_set_word_likelihood
```

[11]:
```
{'this': 0.1081548777849881,
 'hello': 0.021630975556997622,
 'how': 0.15141682889898334,
 'offer': 0.1946787800129786,
 'buy': 0.18386329223447978,
 'thanks': 0.12978585334198572,
 'shipping': 0.2054942677914774,
 'meet': 0.004326195111399524,
 'tomorrow': 0.0006489292667099286,
 'lunch': 0.0,
 'spicy': 0.0,
 'morning': 0.0,
 'evening': 0.0,
 'function': 0.0,
 'party': 0.0}
```

0.0.4 Number of Words in Ham Set and Ham Set Word Likelihood

[12]:
```
ham_set_words = 0
```

[13]:
```
for word in words_in_dict:
    ham_set_words += ham_set[word]
```

[14]:
```
ham_set_word_likelihood = dict()
```

[15]:
```
for word in words_in_dict:
    ham_set_word_likelihood[word] = ham_set[word] / ham_set_words
```

[16]:
```
ham_set_word_likelihood
```

[16]:
```
{'this': 0.24813895781637718,
 'hello': 0.014179369018078695,
 'how': 0.26586316908897556,
 'offer': 0.003544842254519674,
 'buy': 0.003544842254519674,
 'thanks': 0.21269053527118043,
 'shipping': 0.0003544842254519674,
 'meet': 0.10634526763559021,
 'tomorrow': 0.03544842254519674,
 'lunch': 0.02481389578163772,
 'spicy': 0.01772421127259837,
 'morning': 0.014179369018078695,
 'evening': 0.010634526763559022,
 'function': 0.02481389578163772,
 'party': 0.01772421127259837}
```

0.1 Mail to Check

[17]:
```
mail_to_check = "This Offer, Buy."
mail_to_check = mail_to_check.lower()
```

[18]:
```
import re ### I used Regex to seperate words in a mail
words_in_mail = re.split("\s|(?<!\d)[,.](?!\d)", mail_to_check)
```

[19]:
```
words_in_mail
```

[19]:
```
['this', 'offer', '', 'buy', '']
```

0.1.1 Merginal Probability of Words

[20]:
```
merginal_prob = dict()
```

[21]:
```
for word in words_in_dict:
    merginal_prob[word] = spam_set_word_likelihood[word] * spam_prob +␣
↪ham_set_word_likelihood[word] * ham_prob
```

[22]:
```
merginal_prob
```

[22]:
```
{'this': 0.20614373380696044,
 'hello': 0.016414850979754373,
 'how': 0.23152926703197788,
 'offer': 0.06088502358205735,
 'buy': 0.05764037724850771,
 'thanks': 0.187819130692422,
 'shipping': 0.06189641929525959,
 'meet': 0.07573954587833301,
 'tomorrow': 0.025008574561650697,
 'lunch': 0.017369727047146403,
 'spicy': 0.01240694789081886,
 'morning': 0.009925558312655087,
 'evening': 0.007444168734491314,
 'function': 0.017369727047146403,
 'party': 0.01240694789081886}
```

0.1.2 Spam Mail Probability

[23]:
```
spam_mail_prob = 1
for word in words_in_mail:
    if word != "":
        spam_mail_prob *= spam_set_word_likelihood[word] * spam_prob /␣
 ↪merginal_prob[word]
```

[24]:
```
spam_mail_prob
```

[24]:
```
0.14448278095448996
```

0.1.3 Ham Mail Probability

[25]:
```
ham_mail_prob = 1
for word in words_in_mail:
    if word != "":
        ham_mail_prob *= ham_set_word_likelihood[word] * ham_prob /␣
 ↪merginal_prob[word]
```

[26]:
```
ham_mail_prob
```

[26]:
```
0.0014783438891659796
```

0.1.4 Result

[27]:
```
if spam_mail_prob >= ham_mail_prob:
    print('"', mail_to_check, '" is a Spam.')
else:
    print('"', mail_to_check, '" is a Ham.')
```

```
" this offer, buy. " is a Spam.
```

[ ]: