AIMS | African Institute for Mathematical Sciences GHANA

Topic:

NLP

By:

NLP-Group-Members

Book: Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognitio

February 9, 2020

Contents

1 Exercise 4.1

Sentence = "I always like foreign films."?

Class assign by Naive Bayes:

Note: We should neglect the punctuation in our sentence, then the probability for the sentence to belong to positive class is given by equation (1):

$$\mathbb{P}(+|s) = 0.09 \times 0.07 \times 0.29 \times 0.04 \times 0.08 \times 0.5 = 5.8 \times 10^{-6} \times 0.5 \tag{1}$$

and the probability for a sentence to belong to a negative class is given by equation (2):

$$\mathbb{P}(-|s) = 0.16 \times 0.06 \times 0.06 \times 0.15 \times 0.11 \times 0.5 = 9.5 \times 10^{-6} \times 0.5 \tag{2}$$

Then our sentence belongs to negative class.

2 Exercise 4.2

D = fast, couple, shoot, fly

Let us compute the most Likelihood class for D.

$$\mathbb{P}(action) = \frac{3}{5}.$$
$$\mathbb{P}(comedy) = \frac{2}{5}.$$

$$\mathbb{P}(fast|action) = \frac{2+1}{11+7} = \frac{3}{18}.$$

$$\mathbb{P}(couple|action) = \frac{1}{11+7} = \frac{1}{18}.$$

$$\mathbb{P}(shoot|action) = \frac{5}{11+7} = \frac{5}{18}.$$

$$\mathbb{P}(fly|action) = \frac{2}{11+7} = \frac{2}{18}.$$

$$\mathbb{P}(fast|comdey) = \frac{2}{9+7} = \frac{1}{8}.$$

$$\mathbb{P}(couple|comedy) = \frac{3}{16}.$$

$$\mathbb{P}(shoot|comedy) = \frac{1}{16}.$$
$$\mathbb{P}(fly|comedy) = \frac{2}{16}.$$

Now to compute the probability corresponding to each class, we perform the product of probability of all words that belong to the corresponding class.

$$\mathbb{P}(D|action) = 1.7 \times 10^{-4}$$
$$\mathbb{P}(D|comedy) = 7.32 \times 10^{-5}$$

Then we can conclude that our document belong to the class action since it has the maximum probability.

3 Exercise 4.3

3.1 Standard Naive Bayes

$$\mathbb{P}(neg) = \frac{3}{5}.$$
$$\mathbb{P}(pos) = \frac{2}{5}.$$

$$\mathbb{P}(good|pos) = \frac{3+1}{9+3} = \frac{4}{12}.$$

$$\mathbb{P}(great|pos) = \frac{5+1}{9+3} = 0.5.$$

$$\mathbb{P}(poor|pos) = \frac{1+1}{9+3} = \frac{2}{12}.$$

$$\mathbb{P}(good|neg) = \frac{2+1}{14+3} = \frac{3}{17}.$$

$$\mathbb{P}(great|neg) = \frac{2+1}{14+3} = \frac{3}{17}.$$

$$\mathbb{P}(poor|neg) = \frac{10+1}{14+3} = \frac{11}{17}.$$

$$\mathbb{P}(sentence|pos) = \frac{2}{5} \times \left(\frac{1}{5}\right)^2 \times \frac{1}{2} \times \frac{1}{6} = 0.003.$$

$$\mathbb{P}(sentence|neg) = \frac{3}{5} \times \left(\frac{3}{17}\right)^2 \times \frac{3}{17} \times \frac{11}{17} = 0.002.$$

Then the predicted class is positive.

3.2 Binary Naive Bayes

In the case of binary Naive Bayes, One does not care about the frequency of the words inside the documents.

Now the probability becomes:

$$\mathbb{P}(good|pos) = \frac{1+1}{4+3} = \frac{2}{7}.$$

$$\mathbb{P}(great|pos) = \frac{2+1}{4+3} = \frac{3}{7}.$$

$$\mathbb{P}(poor|pos) = \frac{1+1}{4+3} = \frac{2}{7}.$$

$$\mathbb{P}(good|neg) = \frac{2+1}{6+3} = \frac{1}{3}.$$

$$\mathbb{P}(great|neg) = \frac{3+1}{6+3} = \frac{4}{9}.$$

$$\mathbb{P}(poor|neg) = \frac{1+1}{6+3} = \frac{2}{9}.$$

$$\mathbb{P}(sentence|pos) = \frac{2}{5} \times \left(\frac{2}{7}\right)^2 \times \frac{3}{7} = 0.013.$$

$$\mathbb{P}(sentence|neg) = \frac{3}{5} \times \frac{1}{3} \times \frac{4}{9} \times \frac{2}{9} = 0.059.$$

Then the predicted class in negative.