

Machine Learning

Exercise 02

Ciheng Zhang (3473188) zch3183505@gmail.com

Gang Yu (3488292) HansVonCq@gmail.com

Huipanjun Tian (3471607) s Thpjpyl5111217@gmail.com

May 11, 2020

1 Simple Bayes

1.

$$P(A = \text{choosebox1}) = \frac{1}{2}$$

$$P(B = \text{choosebox2}) = \frac{1}{2}$$

$$P(E = \text{chooseapple}|A) = \frac{8}{12} = \frac{2}{3}$$

$$P(E = \text{chooseapple}|B) = \frac{10}{12} = \frac{5}{6}$$

$$P(E = \text{chooseapple}) = P(C|A) * P(A) + P(D|B) * P(B) = \frac{3}{4}$$

$$P(A|E) = \frac{P(E|A)P(A)}{P(E)} = \frac{4}{9}$$

So the probability of choose an apple is $\frac{3}{4}$. The probability of came from box 1 is $\frac{4}{9}$.

2. Spam Classification with Naive Bayes

$$P(A = \text{bag1}) = 0.5$$

$$P(B = \text{bag2}) = 0.5$$

$$P(C = \text{yellow}|A) = 0.2$$

$$P(C|B) = 0.14$$

$$P(D = \text{green}|A) = 0.1$$

$$P(D|B) = 0.2$$

$$P(E) = P(C|B)P(B)P(D|A)P(A) + P(C|A)P(A)P(D|B)P(B) = 0.0135$$

$$P((C|A)|E) = \frac{P(E|(C|A))P(C|A)}{P(E)}$$

$$= P(C|A)P(A)P(D|B)P(B)P(C|A)/P(E) = 0.74$$

So the probability of the yellow one came from bag1994 is 0.74.

2 3. kNN for Text Classification

The answer is in assignment2.ipynb

3 Aufgabe 3

At first people should build a vocabulary, that all the words from dataset. Then if the word appear in the text, it is one. if the text not include the word the value is 0. Then we can build a big vector with 0 and 1 for each text.

Then we can use the angle between the two vector, that made by the input text and train data. This value of the cosine function of this angle are used as the distance function.

The decision rule is we choose the 10 samples ($k=10$) with the smallest angle. And the class of the majority is the class of input text.

The advantage of this methode is easy to implement. But the disadvantage is this vector is too big. that means we need more time to calculate.

4 kNN in High-Dimensional Feature Space

If the dimension of the feature is too big. This means kNN need to calculate more to get the distance. This means more time to Classification. Besides if the feature is too big there is more chance occur the overfitting. People should choose the important feature and give up the useless feature. For example, We can abandon the useless words like "the", "a". We can also choose some easy distance function, Like cosine function.