In [1]:

```
import numpy
def loadDataSet():
    postingList = [['my','dog','has','flea','porblem','help','please'],
                   ['maybe','not','take','him','to','dog','park','stupid'],
                   ['my','dalnation','is','so','cute','I','love','him'],
                   ['stop','postting','ate','my','steak','how','to','stop','him
                   ['mr','licks','ate','my','steak','how','to','stop','him'],
                   ['quit','buying','worthless','dog','food','stupid']]
    classVec = [0,1,0,1,0,1]
    return postingList ,classVec

def createVocabList(dataSet):
    vocabSet = set([])
    for document in dataSet:
        vocabSet = vocabSet | set(document)
        print(vocabSet)
    return list(vocabSet)


#*****************************************************************************
def setofword2vec(voablist,inputSet):
    returnVec = []
    for article in inputSet:
        tmp = [0] * len(voablist)
        for word in article:
            if word in voablist:
                tmp[voablist.index(word)] = 1
            else:
                print("the word:% s is not in my vocabulary" % word)
        returnVec.append(tmp)
        print(returnVec)
    return returnVec
#*****************************************************************************

def bagofwordVec(vocabList,inputSet):
    returnVec = [ ]
    for article in inputSet:
        tmp = [0] * len(vocabList)
        for word in article:
            if word in vocabList:
                tmp[vocabList.index(word)] += 1
            else:
                print(f"the word:% s is not in my vocabulary" % word)
        returnVec.append(tmp)
            # print(returnVec)
    return returnVec

def trainNB(trainMatrix,trainCategory):
    # ":param trainMatrix:"
    numTrainDoc = len(trainMatrix)
    newWords = len(trainMatrix[0])
    pAbusive = sum(trainCategory) / numTrainDoc
    p0num = numpy.ones(newWords)
    p1num = numpy.ones(newWords)
    p0Denom = 2.0
    p1Denom = 2.0
    for i in range(numTrainDoc):
        if trainCategory[i]==1:
```

```
            p1num+=trainMatrix[i]
            p1Denom+=sum(trainMatrix[i])
            print(p1Denom,"p1denom")
        else:
            p0num+=trainMatrix[i]
            p0Denom+=sum(trainMatrix[i])
            print(p0Denom,"p0Deom")
    plvec = numpy.log(p1num/p1Denom)
    p0vec = numpy.log(p0num/p1Denom)
    return pAbusive,plvec,p0vec




def classifNB(vec2classfy,p0vec,p1vec,pclass1):
    p1 = numpy.sum(vec2classfy * p1vec) + numpy.log(pclass1)
    p0 = numpy.sum(vec2classfy * p0vec) + numpy.log(1.0-pclass1)
    if p1>p0:
        return 1
    else:
        return 0

if __name__=='__main__':
    test = [['mr','licks','ate','my','steak','how','food','s']]

    postingList,classVec = loadDataSet()
    VocabList = createVocabList(postingList)
    returnVec =  bagofwordVec(VocabList,postingList)
    pAbusive,p1Vec,p0Vec = trainNB(returnVec,classVec)
    print(pAbusive,p1Vec,p0Vec)
    testVec = bagofwordVec(VocabList,test)
    pclass = classifNB(testVec,p0Vec,p1Vec,pAbusive)
    print(pclass)
```

```
{'my', 'help', 'porblem', 'dog', 'flea', 'has', 'please'}
{'maybe', 'to', 'not', 'dog', 'him', 'park', 'stupid', 'my', 'help', 'please', 't
ake', 'flea', 'has', 'porblem'}
{'maybe', 'so', 'I', 'dog', 'help', 'take', 'please', 'porblem', 'to', 'not', 'da
lnation', 'him', 'park', 'cute', 'stupid', 'love', 'my', 'is', 'has', 'flea'}
{'maybe', 'so', 'steak', 'stop', 'I', 'dog', 'help', 'take', 'please', 'porblem',
'to', 'not', 'dalnation', 'ate', 'postting', 'him', 'park', 'cute', 'stupid', 'lo
ve', 'my', 'is', 'how', 'has', 'flea'}
{'maybe', 'so', 'steak', 'stop', 'I', 'dog', 'help', 'take', 'licks', 'please',
'porblem', 'to', 'not', 'dalnation', 'ate', 'postting', 'him', 'park', 'cute', 's
tupid', 'love', 'my', 'mr', 'is', 'how', 'has', 'flea'}
{'maybe', 'so', 'quit', 'steak', 'stop', 'I', 'dog', 'food', 'worthless', 'help',
'take', 'licks', 'please', 'porblem', 'buying', 'to', 'not', 'dalnation', 'ate',
'postting', 'him', 'park', 'cute', 'stupid', 'love', 'my', 'mr', 'is', 'how', 'ha
s', 'flea'}
9.0 p0Deom
10.0 p1denom
17.0 p0Deom
19.0 p1denom
26.0 p0Deom
25.0 p1denom
0.5 [-2.52572864 -3.21887582 -2.52572864 -2.52572864 -2.12026354 -3.21887582
 -2.12026354 -2.52572864 -2.52572864 -3.21887582 -2.52572864 -3.21887582
 -3.21887582 -3.21887582 -2.52572864 -2.12026354 -2.52572864 -3.21887582
 -2.52572864 -2.52572864 -2.12026354 -2.52572864 -3.21887582 -2.12026354
 -3.21887582 -2.52572864 -3.21887582 -3.21887582 -2.52572864 -3.21887582
 -3.21887582] [-3.21887582 -2.52572864 -3.21887582 -2.52572864 -2.52572864 -2.525
72864
 -2.52572864 -3.21887582 -3.21887582 -2.52572864 -3.21887582 -2.52572864
 -2.52572864 -2.52572864 -3.21887582 -2.52572864 -3.21887582 -2.52572864
 -2.52572864 -3.21887582 -2.12026354 -3.21887582 -2.52572864 -3.21887582
 -2.52572864 -1.83258146 -2.52572864 -2.52572864 -2.52572864 -2.52572864
 -2.52572864]
the word:s is not in my vocabulary
0
```