from math import log

import operator

def calcShannonEnt(dataSet):

numEntries=len(dataSet)

labelCounts={}

for featVec in dataSet:

currentLabel=featVec[-1]

if currentLabel not in labelCounts.keys():

labelCounts[currentLabel]=0

labelCounts[currentLabel]+=1

shannonEnt=0

for key in labelCounts:

prob=float(labelCounts[key])/numEntries

shannonEnt-=prob\*log(prob,2

return shannonEnt

def createDataSet1():

dataSet = [['0', '0', '0',’0’],

['0', '0', '0', '1'],

['0', '0', '1', '0'],

['0', '0', '1', '1'],

['0', '1', '0', '0'],

['0', '1', '0', '1'],

['0', '1', '1', '0'],

['0', '1', '1', '1'],

['1', '0', '0',’0’],

['1', '0', '0',’1’],

['1', '0', '1',’0’],

['1', '0', '1',’1’],

['1', '1', '0',’0’],

['1', '1', '0',’1’],

['1', '1, '1', ’0’],

['1', '1', '1,’1’] ]

labels = ['X1','X2', 'X3']

return dataSet,labels