assignment 3-2

0540170 伏勁松

language：python 2.7

environment：ipython notebook

module：graphlab、math

1、data preprocess

將tranning data 和 testing data 分別載入到excel中，轉換成csv格式的數據。

對於missing data，用該feature和status對應的average值替換。

最終得到兩個文件 training_data.csv 和 testing_data.csv

2、導入文件

```
training_data = graphlab.SFrame('training_data.csv')
testing_data = graphlab.SFrame('testing_data.csv')
```

3、使用exponential distribution

4、對training data進行status分組

```
class1 = training_data[training_data['STATUS'] == 'normal']
class2 = training_data[training_data['STATUS'] == 'settler']
class3 = training_data[training_data['STATUS'] == 'overmean']
class4 = training_data[training_data['STATUS'] == 'solids']
class5 = training_data[training_data['STATUS'] == 'low']
class6 = training_data[training_data['STATUS'] == 'storm']
clses = [class1,class2,cls3,class4,class5,class6]
print len(cls1),len(cls2),len(cls3),len(cls4),len(cls5),len(cls6)
```

```
314 6 108 4 69 3
```

5、運用3-1的function計算probability

```
clist = ['class1','class2','class3','class4','class5','class6']
probailities = []
for test in testing_data:
    k = 0
    prolist = []
    for c in clist:
        pro = 1
        i = 0
#         print c
#         print k
        clspara = classparameters[k]
#         print clspara
        for f in features:
            pro = pro * getExpDis(test[f],float(1)/(clspara[i].get('mean')))
            i = i+1
        pro = pro * float(len(locals()[c])) / len(training_data)
        prolist.append(pro)
        k = k+1
    probailities.append({test['DATE']:prolist})
print probailities
```

6、找出每個query中probaility最大的status

```
for pp in probailities:
    date = pp.keys()
    pl = pp.values()[0]
#     print pl
    status = clist[pl.index(max(pl))]
    print date,'屬於' ,status
# probailities[12].values()[0]
```

7、最終結果

```
['D-1/4/90'] 屬於 class1
['D-2/4/90'] 屬於 class1
['D-3/4/90'] 屬於 class1
['D-4/4/90'] 屬於 class1
['D-5/4/90'] 屬於 class1
['D-6/4/90'] 屬於 class1
['D-8/4/90'] 屬於 class1
['D-9/4/90'] 屬於 class1
['D-10/4/90'] 屬於 class1
['D-11/4/90'] 屬於 class1
['D-13/4/90'] 屬於 class1
['D-16/4/90'] 屬於 class1
['D-17/4/90'] 屬於 class5
['D-18/4/90'] 屬於 class1
['D-19/4/90'] 屬於 class1
['D-20/4/90'] 屬於 class1
['D-22/4/90'] 屬於 class3
['D-23/4/90'] 屬於 class1
['D-24/4/90'] 屬於 class1
['D-25/4/90'] 屬於 class1
['D-26/4/90'] 屬於 class2
['D-27/4/90'] 屬於 class1
['D-29/4/90'] 屬於 class2
```