Weekly Report of Research Work
WR-ABS-TEMP-2015A-No.002

汤吉(Ji TANG)
Number: WR-ABS-TEMP-2015A, E-mail: tangji08@hotmail.com

Date: 23/11/2015 - 29/11/2015

November 29, 2015

Contents

1 Work

This week, I have done three main things:

1. Read the book ”Machine Learning In Action” about kNN
2. Watch the open classes online of Stanford about Machine Learning
3. Purchase a domain name ”yejojo.com” for our VPS and many things related to it

2 A Simple example in the book

At first, we’ll create a kNN.py

```
from numpy import *
import operator

def createDataSet():
    group = array([[1.0,1.1],[1.0,1.0],[0,0],[0,0.1]])
    labels = ['A','A','B','B']
    return group,labels
```

Then we want to put the kNN classification algorithm into action:

```
def classify0(inX, dataSet, labels, k):
    dataSetSize = dataSet.shape[0]
    diffMat = tile(inX, (dataSetSize,1)) - dataSet
    sqDiffMat = diffMat**2
    sqDistances = sqDiffMat.sum(axis=1)
    distances = sqDistances**0.5
    sortedDistIndicies = distances.argsort()
    classCount={}
    for i in range(k):
        voteIlabel = labels[sortedDistIndicies[i]]
        classCount[voteIlabel] = classCount.get(voteIlabel,0) + 1
    sortedClassCount = sorted(classCount.iteritems(), key = operator.itemgetter(1)
        , reverse = True)
    return sortedClassCount[0][0]
```

In the library ”numpy” and ”operator”, we have

1. tile (* , (* , *)): for a=[[1,2,3], [4,5,6]]
   tile(a,[2,1]) = array([[1, 2, 3], [4, 5, 6], [1, 2, 3], [4, 5, 6]])
   tile(a,[1,2]) = array([[1, 2, 3, 1, 2, 3], [4, 5, 6, 4, 5, 6]])

2. *.sum: *.(axis = 0) means the sum of each line and *.(axis = 1) means the sum of each row

3. *.argsort: Return the index of *

4. *.get: Get the value of *.(If it doesn’t exist, let it become 0.)

5. sorted(* ,* ,*): Sort every value in * according to *, and the sequence will be reverse or not(reverse = True).

And finally, we could test this simple kNN method to classify the 2D-points.

```
import kNN
group, labels = kNN.creatDataSet()
kNN.classify0([0,0.2], group, labels, 3)
```

And we get the result: ”B”

3 The open classes online of Stanford

Then I have watched two open classes online of Stanford about machine learning. The first class:[1] is about ”Linear Regression”, ”Gradient Descent” and ”Normal Equations”. These all things I have learned in the class optimization this semester, but I’d like to say that I haven’t totally grasped them. After the Andrew Ng’s class, I think there is a big progress for me in the field of machine learning. And the second class:[2] is talking about K-means Clustering Algorithm, which I have seen in the books of machine learning.

The next week, I’m trying to programme a simple programme by using the language Python about K-means Algorithm, then I’ll compare it with the programme in the book and find out my defaults.

4 Parsing the domain name

Last week, we have purchase a VPS in American and set up a VPN service for better researching information on the Internet. But I think this waste a lot of resources because we use a whole VPS just for VPN service. Then I decided to make full use of it:I purchase a domain name ”yejojo.com” and resolve it to our VPS IP. After that, I have downloaded a formwork of html and uploading it to our VPS.(Figure 1)

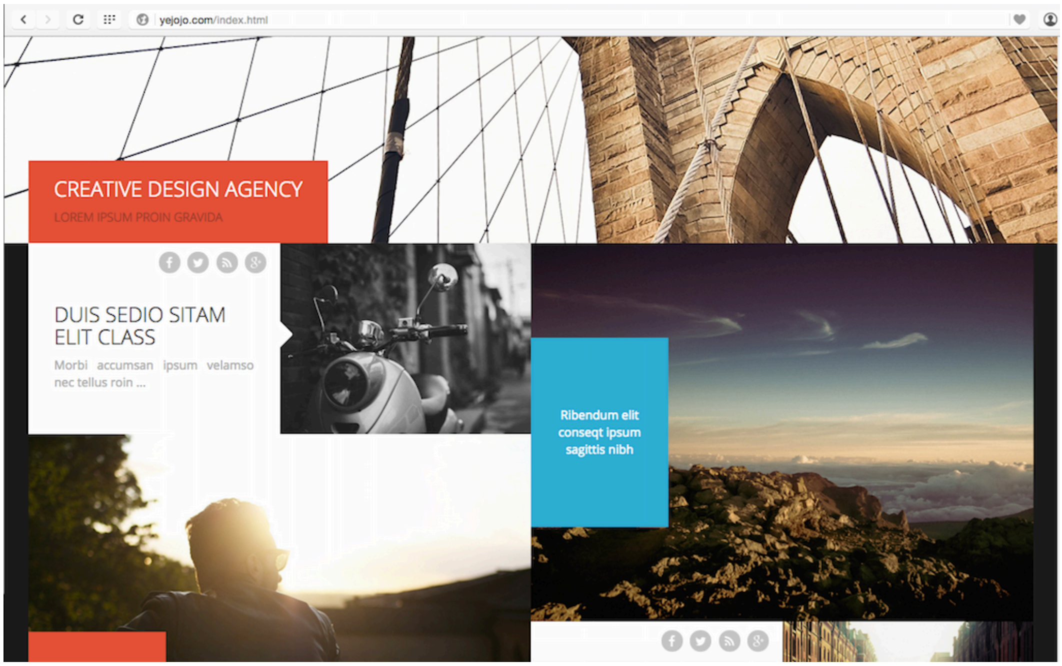
yejojo.com/index.html
CREATIVE DESIGN AGENCY
LOREM IPSUM PROIN GRAVIDA
DUIS SEDIO SITAM ELIT CLASS
Morbi accumsan ipsum velamso nec tellus roin ...
Ribendum elit conseqt ipsum sagittis nibh

Figure 1: The screenshot of yejojo.com

References

[1] Andrew Ng. An application of supervised learning - autonomous deriving. https://see.stanford.edu/Course/CS229/54.

[2] Andrew Ng. The concept of unsupervised learning, k-means clustering algorithm, k-means algorithm. https://see.stanford.edu/Course/CS229/43.