CS 432/532: Web Science

Spring 2017

Assignment 10

Michael Micros

Instructor: Michael L. Nelson

Old Dominon University

Norfolk, Virginia

May 1, 2017

Problem 1: Computing kNN for Dr. Nelson's blogs

Having substituted the euclidean distance with the cosine, the Knn function provided in the "Programming Collective Intelligence" textbook was used on the blogdata collected in Assignment 8. The blog data are located in "blogdata.txt".

```python
f = open("blogdata.txt", "r")

for line in f:
  a = line.strip('\n').split('\t');
  b = a.pop(0)
  vecs[b] = a

print len(vecs)

fm = 'F-Measure'
ws = 'Web Science and Digital Libraries Research Group'

a = vecs[fm]
temp = vecs.values()
temp.pop(vecs.keys().index(fm))

a = knnestimate(temp,a,k=5)


k = [1, 2, 5, 10, 20]
print "\n------F-Measure kNN------"
for i in k:
  print "\n---k = "+str(i)
  for j in range(i):
    b = a[j][1]
    print vecs.keys()[b]

a = vecs[ws]
temp = vecs.values()
temp.pop(vecs.keys().index(ws))

a = knnestimate(temp,a,k=5)


k = [1, 2, 5, 10, 20]
print "\n-----WS-DL kNN-----"
for i in k:
  print "\n---k = "+str(i)
  for j in range(i):
    b = a[j][1]
    print vecs.keys()[b]
```

Figure 1: Code for calculating kNN for F-Measure and WS-DL.

Figure 2: kNN results for http://f-measure.blogspot.com/.

Figure 3: kNN results for http://ws-dl.blogspot.com/.

Problem 2: LIBSVM