Assignment 10

Part 1:

The codes I used for part 1 and 2 were used from a combined effort of Kevin Clemmons and Gabriel Marquez

I was able to read my blog data from the previous past 2 assignments, assignment 8; and it was able to print out data with a cosine distance of the 500 dimension, I used knnestimate() to be able to locate the nearest neighbors for the two links you provided, http://f-measure.blogspot.com/and http://ws-dl.blogspot.com/.

```
def getdistances(data,vec1):
    distancelist=[]
    for i in range(len(data)):
        vec2=data[i]
        distancelist.append((Cosine(vec1,vec2),i))
    distancelist.sort( )
    return distancelist

def knnestimate(data,vec1,k=3):
    dlist=getdistances(data,vec1)
    avg=0.0
    for i in range(k):
        idx=dlist[i][1]
        avg+= idx
    avg=avg/k
    return avg
```

```
Z:\cs432\Assignment10>python kMeasure.py
44.0
28.0
29.0
27.2
24.45
50.0
43.5
29.0
27.4
29.1

Z:\cs432\Assignment10>
```

	K=1	K=2	K=5	K=10	K=20
http://f-measure.blogspot.com/	44	28	29	27.2	24.45
http://ws-dl.blogspot.com/	50	43.5	29	27.4	29.1

Part 2:

I used the LIBSVM to rerun assignment 9 question 2 to be able to classify my list, I reran the program 4 times being that I had 4 categories. Using the script I manually assigned 50 classifications and it made estimations on the others; I ran them on both the true and negated sides. According to my results Actors had the highest percentage for correctness.

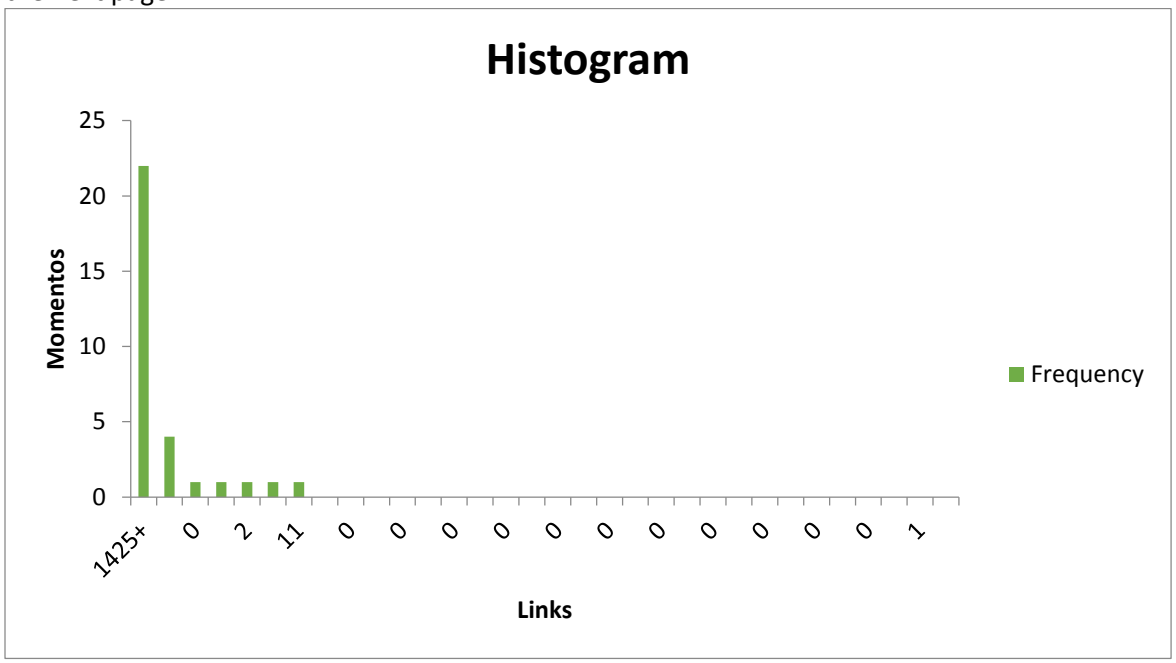
In order to make my program for my LIBSVM to run I incorporated snipptes from LIBSVM tutorials and websites and adjusted it to my needs.

http://rstudio-pubs-static.s3.amazonaws.com/22171_ac34b133d26c4affb178031a9407d1ce.html#support-vector-
machinehttps://www.youtube.com/watch?v=iu9rrCbSqgM&list=PLqS2sO7F2t3XKyCABKHxWO_x56X_DGRc-

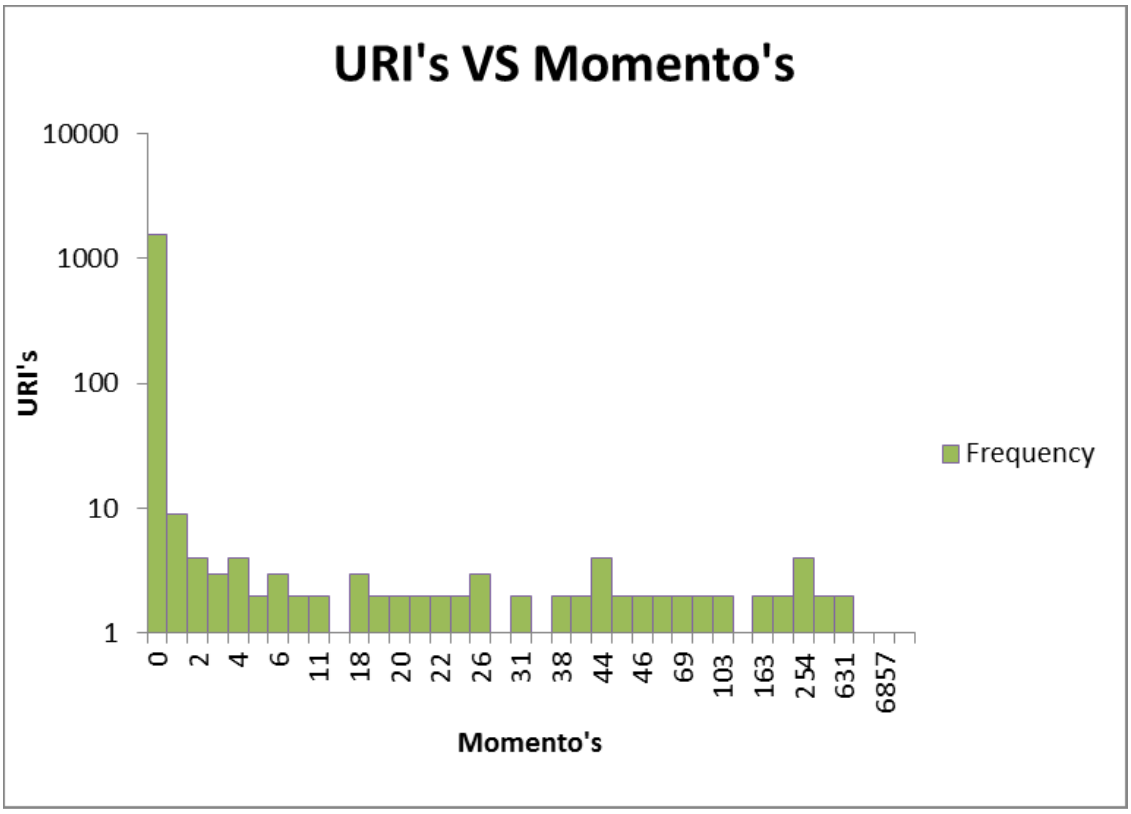
https://github.com/cjlin1/libsvm/blob/master/python/svm.py

	Cross Validation %	Correct %
Movie	65%	54%
Directors	76%	59%
Entertainment	47%	35%
Actors	83%	86%

Part 3(Extra Credit): New Histogram
For this section I used my twitter links from assignment 2 and created a histogram to make a comparison of the change over the past few months, my chart has changed drastically as you can see on the next page.



Old Histogram:


URI's VS Momento's
10000
1000
100
10
1
URI's
0
2
4
6
11
18
20
22
26
31
38
44
46
69
103
163
254
631
6857
Momento's
Frequency