SEng 474

Assignment 3

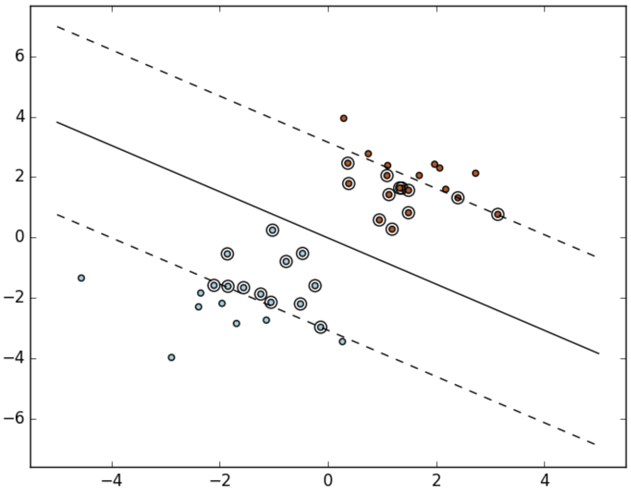
Jakob Roberts
v00484900

1

a) 3 Support vectors, because of data threshold.

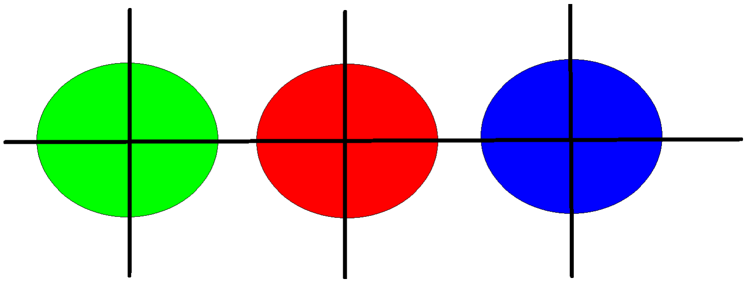
b) 24 Support Vectors

c)

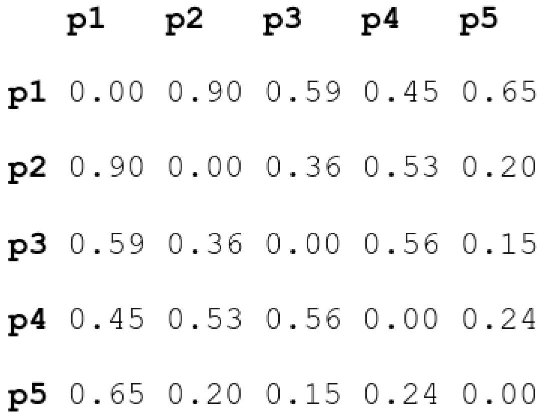


2

```
Consider a data set that consists of three circular clusters, that are identical
in terms of the number and distribution of points, and whose centers lie on
a line and are located such that the center of the middle cluster is equally
distant from the other two. Bisecting K-means would always split the middle
cluster during its first iteration, and thus, could never produce the correct
set of clusters.
```



3



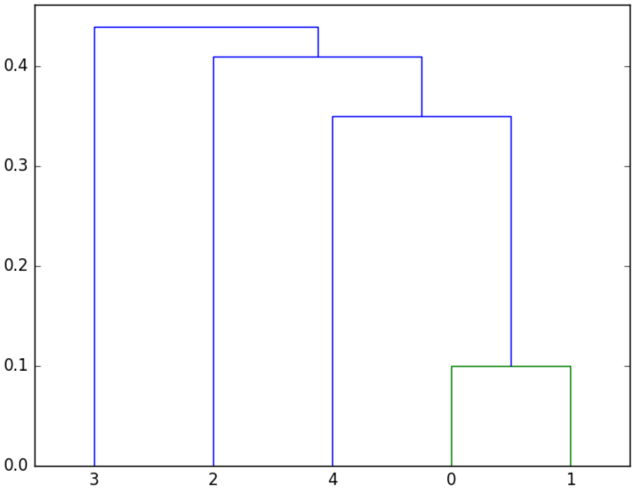
	p1	p2	p3	p4	p5
p1	0.00	0.90	0.59	0.45	0.65
p2	0.90	0.00	0.36	0.53	0.20
p3	0.59	0.36	0.00	0.56	0.15
p4	0.45	0.53	0.56	0.00	0.24
p5	0.65	0.20	0.15	0.24	0.00

MIN:

```
import scipy.spatial.distance as ssd
import hcluster
import pylab
import matplotlib.pyplot as plt

SimMatrix = [[0.00, 0.90, 0.59, 0.45, 0.65],
[0.90, 0.00, 0.36, 0.53, 0.20],
[0.59, 0.36, 0.00, 0.56, 0.15],
[0.45, 0.53, 0.56, 0.00, 0.24],
[0.65, 0.20, 0.15, 0.24, 0.00]]

distVec = ssd.squareform(SimMatrix)
linkage = hcluster.linkage(1 - distVec)
dendro  = hcluster.dendrogram(linkage,distance_sort=”ascending”)
plt.show()
```

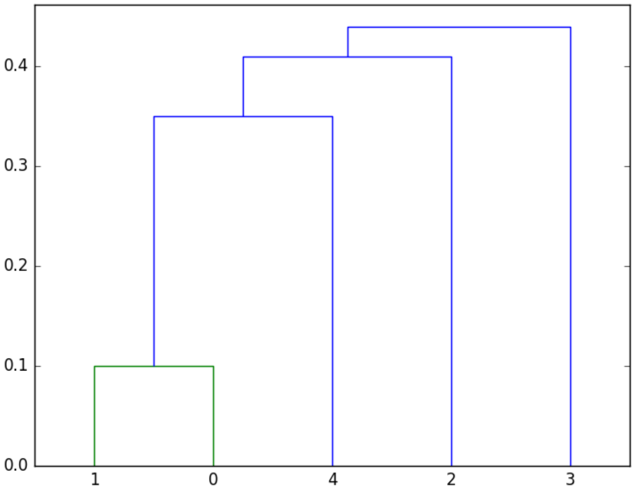


MAX:

```
import scipy.spatial.distance as ssd
import hcluster
import pylab
import matplotlib.pyplot as plt

SimMatrix = [[0.00, 0.90, 0.59, 0.45, 0.65],
[0.90, 0.00, 0.36, 0.53, 0.20],
[0.59, 0.36, 0.00, 0.56, 0.15],
[0.45, 0.53, 0.56, 0.00, 0.24],
[0.65, 0.20, 0.15, 0.24, 0.00]]

distVec = ssd.squareform(SimMatrix)
linkage = hcluster.linkage(1 - distVec)
dendro  = hcluster.dendrogram(linkage,distance_sort=”descending”)
plt.show()
```


0.4
0.3
0.2
0.1
0.0
1
0
4
2
3