12/7/2016 Clustering

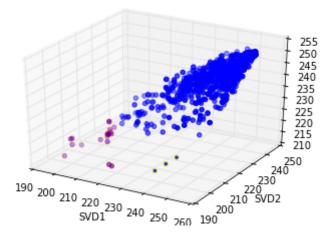
```
In [17]: %matplotlib inline

import mltools.cluster as cluster
import mltools as ml
import numpy as np
import matplotlib.pyplot as plt
import mltools.dtree as dtree
from scipy import linalg

Xtr = np.genfromtxt("data/X_train.txt", delimiter=None)
Ytr = np.genfromtxt("data/Y_train.txt", delimiter=None)
```

```
In [14]: #z, join = cluster.agglomerative(Xtr[:100,:3], 5)
```

```
In [22]: from mpl_toolkits.mplot3d import Axes3D
         U, s, V = linalg.svd( Xtr[:10000,:], full_matrices=False )
         \#Sig = mat(eye(S)*s[:S])
         #tak out columns you don't need
         newdata = Xtr[:1000,:3]
         z, join = cluster.agglomerative(newdata, 3)
         # this line is used to retrieve dataset
         \#~ new = U[:,:2]*Sig*V[:2,:]
         fig = plt.figure()
         ax = fig.add subplot(111, projection='3d')
         colors = ['blue','red','yellow', 'black', 'green']
         c= [colors[int(clusterNo)] for clusterNo in z ];
         ax.scatter(newdata[:,0],newdata[:,1],newdata[:,2], color= c)
         #for i in xrange(Xtr.shape[0]):
              ax.scatter(newdata[i,0],newdata[i,1])
         plt.xlabel('SVD1')
         plt.ylabel('SVD2')
         plt.savefig('svdWithClustering.png')
```



12/7/2016 Clustering