In [1]:

```
import csv as csv
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn import svm
from scipy import linalg
from sklearn import mixture
from sklearn.decomposition import PCA
from sklearn.preprocessing import scale
```

In [2]:

```
# Open up the csv file in to a Python object
data = pd.read_csv('ass3_data.txt',header = -1,sep=' ')
data = data.fillna(0)
```

In [3]:

```
s2 = linalg.svd(data, compute_uv=False)
# print(s2)

X = np.zeros(s2.size)
for i in range(s2.size):
    X[i] = i+1
plt.plot(X,s2,'ro')
plt.grid(True)
plt.show()
```

In [3]:

```
print("Running PCA");
#Run PCA to compress data
pca = PCA(n_components=32)
pca.fit(data)
transformed_data = pca.transform(data)
print(pca.n_components_)
# print("Running GMM");
# gmm = mixture.GMM(n_components=clusters)
# gmm.fit(transformed_data)
# result = gmm.predict(transformed_data)
# np.savetxt('sub.csv', result,newline=',')
print("done");
```

```
Running PCA
32
done
```

In [19]:

```

```

In [ ]:

```
```

In [4]:

```
clusters = 14
print("Running GMM");
gmm = mixture.GMM(n_components=clusters,covariance_type='full')
gmm.fit(transformed_data)
result = gmm.predict(transformed_data)

np.savetxt('sub1.csv', result,newline=',')
print("done");
print(gmm.converged_)
```

```
Running GMM
done
True
```

In [17]:

```
result /= 1
np.savetxt('sub1.csv', result,newline=',')
print("done");
print(gmm.converged_)
```

```
done
True
```

Experiment Log

Dimension = 10

- Accuracy increases by increasing the cluster no from 10 to 12 but decreased for 15 ### Dimension = 55 (90 % variance)
- c = 12 accuracy = 55
- c = 15 accuracy = 58
- c = 18 accuracy = 58.9
- c = 16 accuracy = 57
- c = 21 accuracy = 58.2
- c = accuracy = 58 ### Dimension = 78 (95 % variance) Scaled data
- c = 15 accuracy = 50 ## Scaling is not a good thing ### Dimension = 78 (95 % variance)
- c = 15 accuracy = 56.9 ### Dimension = 32 (80 % variance)
- c = 15 accuracy = 56.9
- c = 18 accuracy = 56.9 ### Dimension = 32 (80 % variance) cov = tied
- c = 6 accuracy =
- c = 12 accuracy =
- c = 15 accuracy = 56.9
- c = 18 accuracy = 56.