HW2 - ml4786

In [72]:
```python
%matplotlib inline

import numpy as np
from sklearn.cluster import KMeans
from scipy.spatial.distance import cdist
import matplotlib.pyplot as plt
import os
print("Place data files in '"+os.getcwd()+"/'")
```

Place data files in '/home/neil/Projects/ML4786/HW 2 - JAC/'

In [73]:
```python
def csv_to_matrix(csv_file):
    return np.loadtxt(open(csv_file),delimiter=",")
```

In [74]:
```python
def my_kmeans(X,us): #returns the cluster arrays
    km = KMeans(n_clusters=2, init=us, n_init=1)
    km.fit(X)
    classifications = km.predict(X)
    C1 = []
    C2 = []
    for c,point in zip(classifications, K1):
        if c==0:
            C1.append(point)
        else:
            C2.append(point)
    return np.array(C1), np.array(C2)
```

In [80]:
```python
def plot_2cluster(C1, C2, US):
    plt.scatter(C1[:,0],C1[:,1], c='red',  alpha=1, marker="o", label="C_1")
    plt.scatter(C2[:,0],C2[:,1], c='blue', alpha=1, marker="o", label="C_2")
    plt.scatter(US[:,0],US[:,1], c='green', alpha=1, marker="+", label="Centroids")
    plt.legend(loc='upper right')
    plt.show()
```

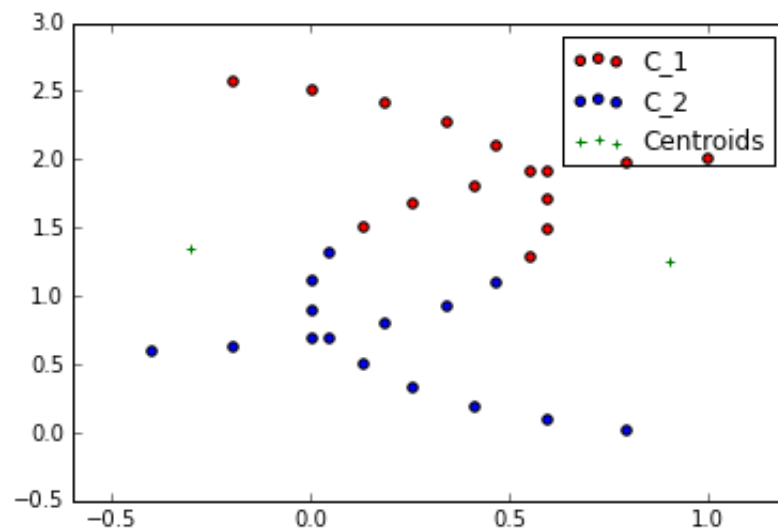# Kmeans_I

```
In [81]:  # Get data
          K1 = csv_to_matrix('XkmeansI.csv')
          u1 = [0.9, 1.25]
          u2 = [-0.29999999999999993, 1.35]
          us = np.array([u1, u2])
          # Run kmeans
          K1_C1, K1_C2 = my_kmeans(K1,us)
          # Plot
          plot_2cluster(K1_C1, K1_C2, us)
```
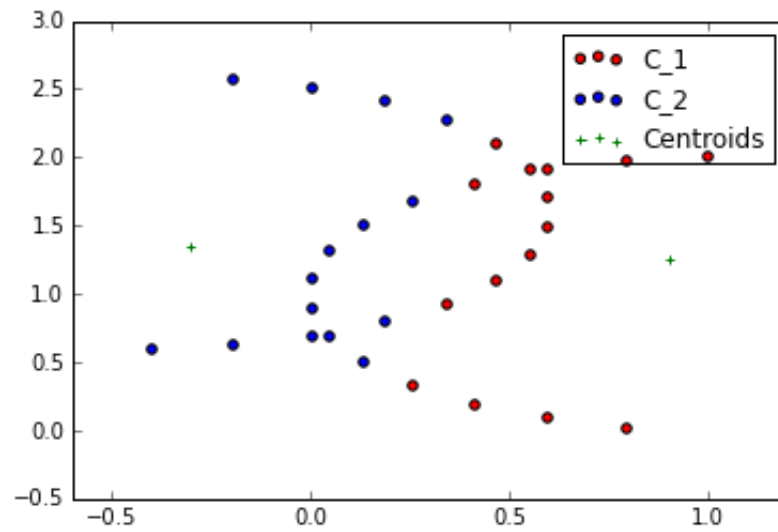

Kmeans_2

```
In [83]: # Get data
         K2 = csv_to_matrix('XkmeansII.csv')
         u1 = [0.9, 1.25]
         u2 = [-0.29999999999999993, 1.35]
         us = np.array([u1, u2])
         # Run kmeans
         K2_C1, K2_C2 = my_kmeans(K2,us)
         # Plot
         plot_2cluster(K2_C1, K2_C2, us)
```



```
In [ ]:
```