# Pattern and Speech Recognition WS2015-16
# Exercise 5

Atanas Poibrenski(2554135), Marimuthu Kalimuthu(2557695), Furkat Kochkarov(2557017)

December 4, 2015

**Gaussian Mixture Models**

## 1 Data Preparation

**Ex-1** Done.

**Ex-2** Loaded data into 677970x50 matrix and removed the first column(which is the useless dimension). (See 'data_preparation.m')

**Ex-3** We used *pca* function of matlab and selected the first dimension of the projected data. (See data_preparation.m)

**Ex-4** We plot the distribution using histogram with 200 bins to get a clear idea of the distribution of the data. The cluster size is 5 because the histogram resembles five separate groups of points.

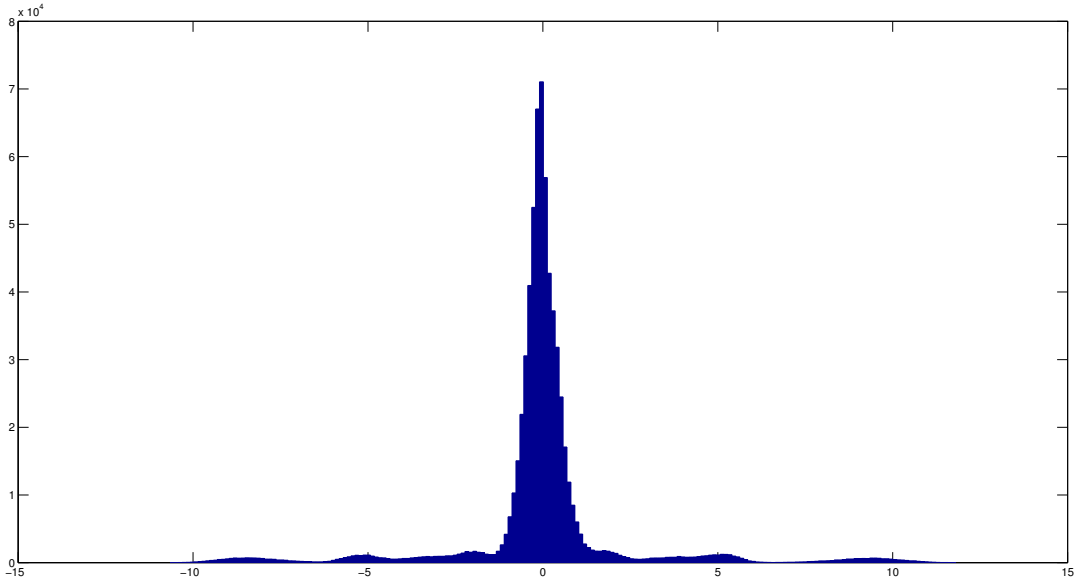


Figure 1: Data distribution - histogram with 200 bins

# 2 Clustering uing K-Means

## 2.1 Cluster Association

**Ex-5** See 'rmse.m'

**Ex-6** See 'associate.m'

## 2.2 Compute Means

**Ex-7** It should be (k x N) since we want to compute **k** amount of *means* and each *mean* will be of the dimension of the data which is **N**.

**Ex-8** See 'compute_means.m'

## 2.3 Initialization

**Ex-9** We set the initial **k** means as random **k** points from the data. This performs better than random initialization. When we used random initializations, some clusters remain unassigned to data points which requires additional steps to deal with.

## 2.4 K-Means

**Ex-10** See 'kmeans_.m'

**Ex-11** cluster means plot

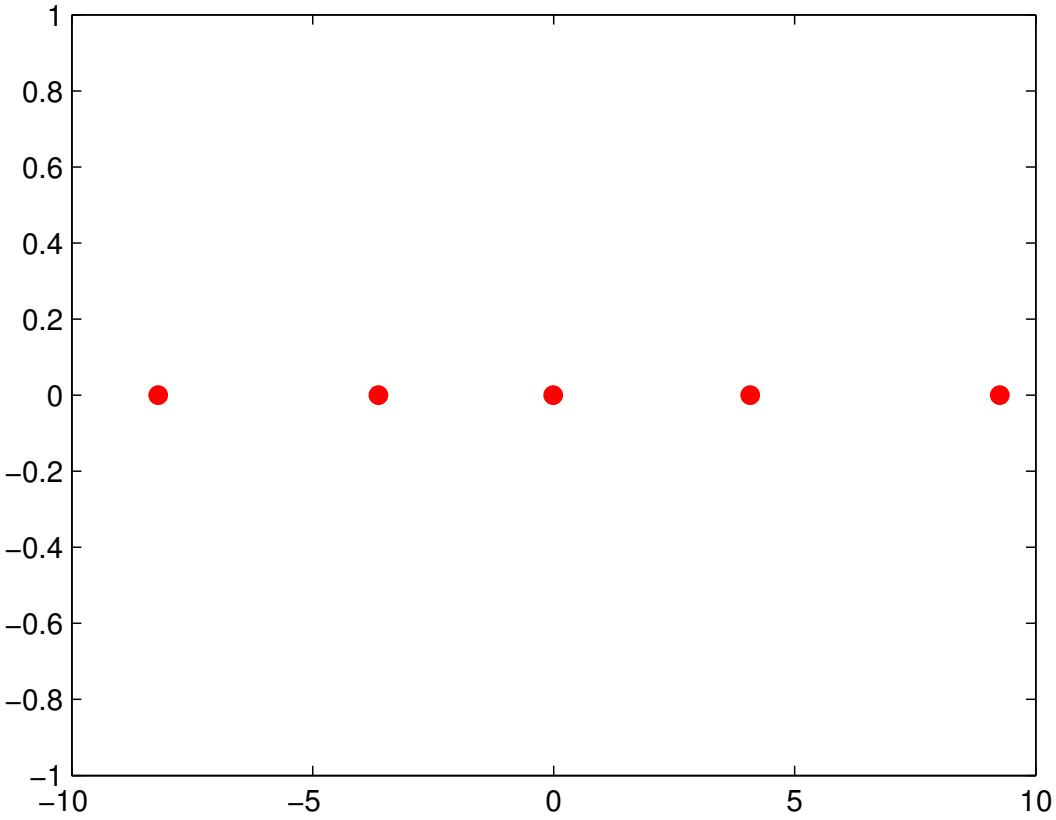


Figure 2: means of all clusters

## 3 GMM initialization

**Ex-12** Means: [-0.0086 -3.6375 -8.2074 9.2524 4.0791 ]

Covs: [0.2669 1.5800 0.9207 0.8581 1.3088]

Weights: [ 0.8650 0.0502 0.0209 0.0198 0.0438 ]

(Also see 'gmm_init.m')

**Ex-13** We have defined 5 Gaussians with separate weights which is the definition of Gaussian Mixture Model. Instead of using EM algorithm, we used the final clusters from 'kmeans' algorithm.

**Ex-14** (**Bonus**) GMM-Plot

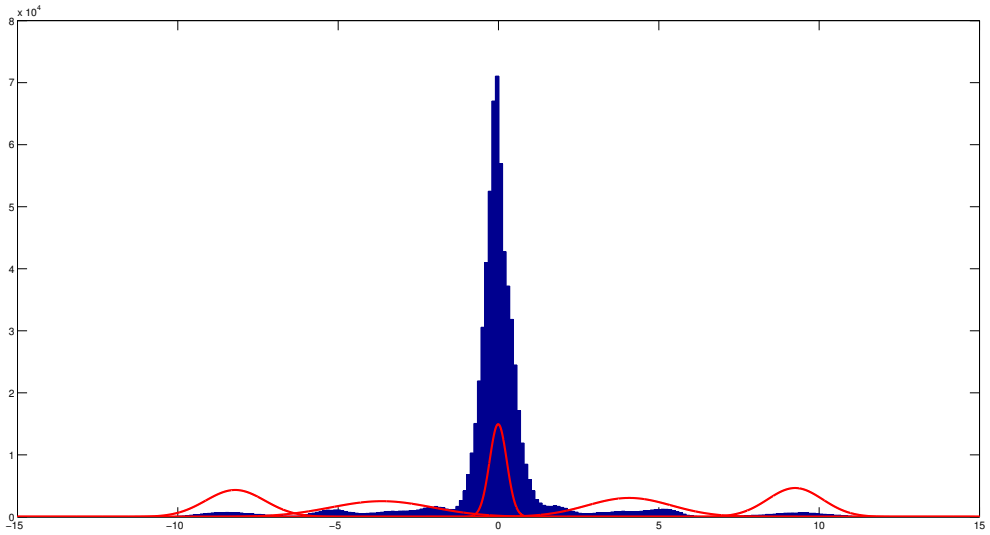


Figure 3: Overlay of histogram and Gaussians

# 4 Application to the real data set

**Ex-15** Done. See all code!