**Problem Set 2**

**Show that the K-means procedure can be viewed as a special case of the EM algorithm**

**applied to an appropriate mixture of Gaussian densities model.**

**Sources:**

<https://alliance.seas.upenn.edu/~cis520/wiki/index.php?n=Lectures.EM>

K-Means can be viewed as a special case of clustering with a mixture of Gaussians where all variances are equal under the assumption that the clusters are spherical.

**You have to communicate a signal in a language that has 3 symbols A, B and C. The**

**probability of observing A is 50% while that of observing B and C is 25% each. Design an**

**appropriate encoding for this language. What is the entropy of this signal in bits?**

Using variable width encoding, assigning:

A = 1

B = 2

C = 2

Entropy, S = (0.5 + 0.25\*2 + 0.25\*2) = 1.5

**Plot the direction of the first and second PCA components in the figures given**

Chart, line chart

Description automatically generated