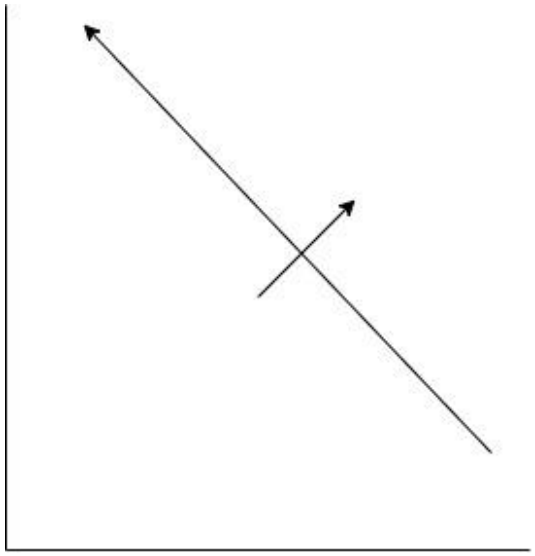
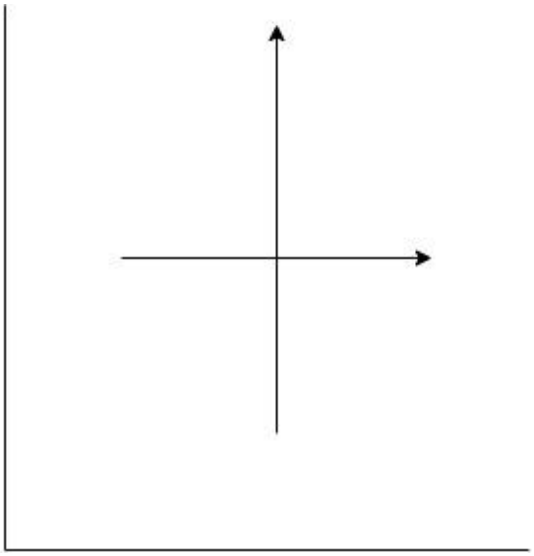
1. A language with only 3 symbols where A is 50% B and C are 25% each. A can be represenSted by a single bit and B and C both occur half as often so they will need two bits.

$$1 * (0.5) + 2 * (0.25) + 2 * (0.25) = 1.5\ total\ bits$$

2. K Means clustering can be viewed as a special case of EM algorithm when applied to the appropriate mixture of Gaussian densities model when the variance is not considered.

3.  

4. Which Clustering Methods at K=2?
    a. Hierarchical with Single Link because the separation between the clusters is well defined. K means would not work well unless the centroids were very far away from these clusters.
    b. K Means, because you could have the centroid at the center of each spiral and it would work well. I would expect that Hierarchical would not work well due to how close the points are at the cluster boundaries. EM I would also expect it to work well.
    c. K Means would not work because there are mixed up points and for those instances you would need a soft clustering algorithm such as EM