# Homework 6

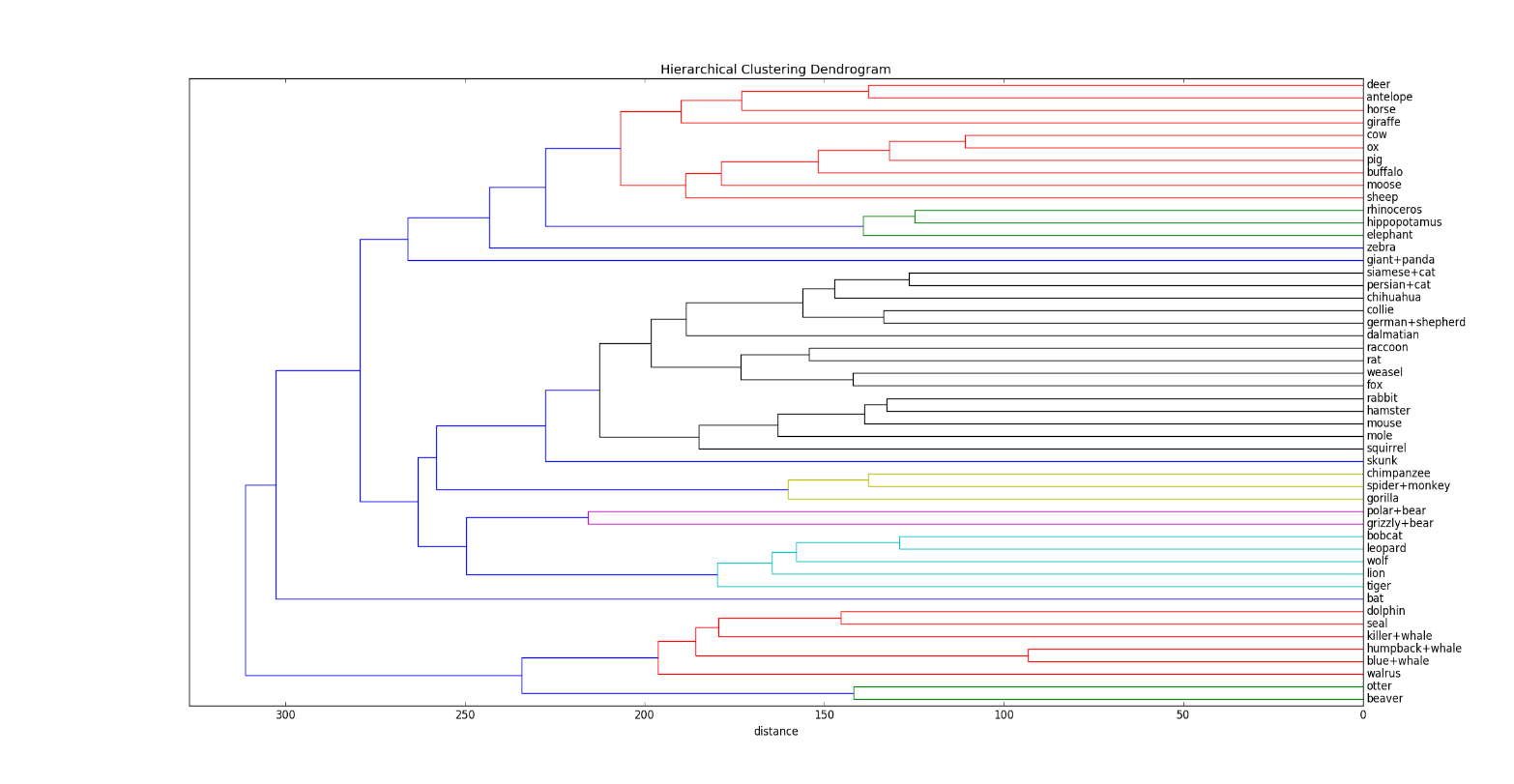
**Problem 1:**

The following is the list of k-means clusters:

|  |  |
| --- | --- |
| **Cluster No.** | **Animals** |
|  | Dalmatian, Persian cat, German shepherd, Siamese cat, Chihuahua, collie |
|  | Moose, ox, sheep, buffalo, giant panda, pig, cow |
|  | Spider monkey, gorilla, chimpanzee |
|  | Killer whale, blue whale, humpback whale, seal, otter, walrus, dolphin |
|  | Polar bear, grizzly bear |
|  | Tiger, leopard, fox, wolf, bobcat, lion |
|  | hippopotamus, elephant, rhinoceros |
|  | beaver, skunk, mole, hamster, squirrel, rabbit, rat, weasel, mouse, raccoon |
|  | antelope, horse, giraffe, zebra, deer |
|  | bat |

The clusters makes sense since animals with similar features are in the same cluster.

Dendron Diagram:



The hierarchical clusters makes sense since similar animals are grouped first.

**Problem 2:**

1. Consider a cluster C. Let n be the number of points x in the cluster, so that:

To find the optimal µ, we need to minimize the function with respect to µ:

Taking partial derivative of f with respect to (ie we get,

Hence for all values of i

Thus we can write ie, µ is the mean of points in C.

Hence proved.

1. Consider three points in a cluster x(1) = 0, x(2) = 4, x(3) = 5 in one dimension. According to k means the optimal center of the cluster should be = 3. If l1 distance is chosen, then. But if then f(x) = 5. Hence l1 distance cannot be used for k means.

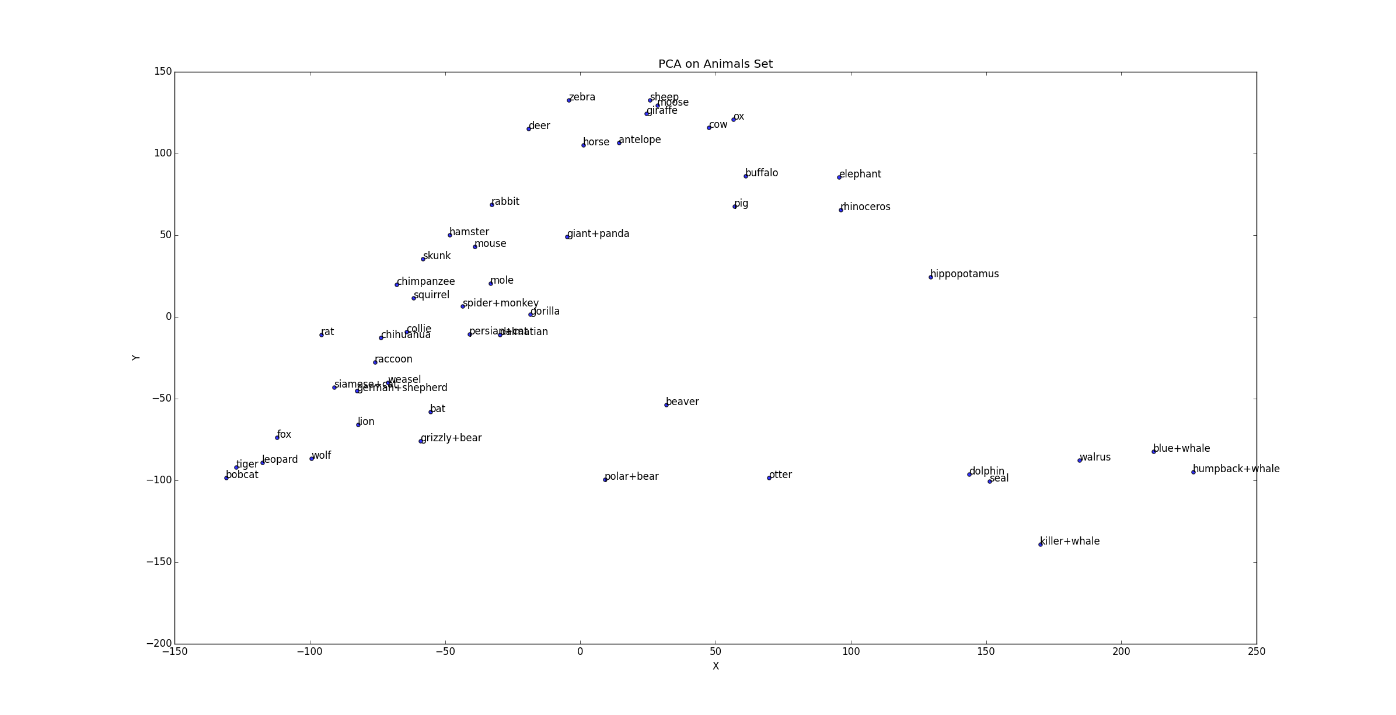
For (R1, l1), the optimal center can be characterized as the median of points in the cluster.

**Problem 3:**

1. The optimal k-means solution is : [-9, 0, 9]
2. Let the initialization of the centers be [0, 8, 10]. In this case, the final answer will have centers as [-6, 8, 10] and clusters as [{-10, -8, 0}, {8}, {10}] which is not the optimal solution.

**Problem 4:**

Visualization of Animal set in 2d by performing PCA:

The graph makes sense since similar animals are grouped nearby.

**Problem 5:**

1. The following are the dimensions of the given matrices in the form (rows, columns)

* Dim(U) = (p, 2)
* Dim(UT) = (2, p)
* Dim(UUT) = (p, p)
* Dim(u1u1T) = (p, p)

1. Projection 1 and 3 are the same. They represent the projection of x in a 2d subspace defined by (u1, u2)

Projection 2 and 4 are the same. The represent projection of x on 2 directions defined by u1 and u2. It shows the projection of x as a p dimensional vector.