CHUNYUAN LI

JIARUI XING

# CSE 517A Machine Learning THW3

## (30 points) Parameter Learning for Gaussian Processes (GPs)

### Note:

Then:

Therefore:

i.e.

## (25 points) K-means Clustering

### The two conditions are equivalent to each other. Proof:

### (i)(ii):

If assignment do not change, i.e. in do not change, will not change, i.e. (ii) holds;

(ii)(i)

If cluster centers do not change, i.e. in

do not change, will not change, i.e. (i) holds.

### K-means always converge. Proof:

(i) The object function will always decrease after updating assignments (until converge, when cluster centers are fixed):

For any data point , since are assigned with the closet cluster center, changes only if updated can decrease . Otherwise, does not change and so as . Therefore, the objective function will either decrease or keep unchanged.

(ii) The object function will always decrease after updating cluster centers (until converge, when assignments are fixed):

Here we are going to show is the global minima for .

To find the optimal point, let:

And:

Which means the objective function is convex (with fixed assignments) and is the optimal solution. Therefore, the object function will always decrease after updating cluster centers.

(iii) From (i) and (ii) we know that the objective function will always decease in both two steps of k-means until assignments or cluster centers does not change, the algorithm must converge to some local minima point.

### It is possible to generate empty cluster. Eg. Empty cluster exists when the user defined cluster number greater than the number of data points.

If the number of clusters is wrong, it might generate empty cluster as well. For an example, in the following figure, we set the number of clusters to be 3. The shapes of the points are meant to show that there are actually 2 natural clusters. But we do not know the natural number of clusters before running a K-means algorithm, so currently assume we have chosen k to be 3.

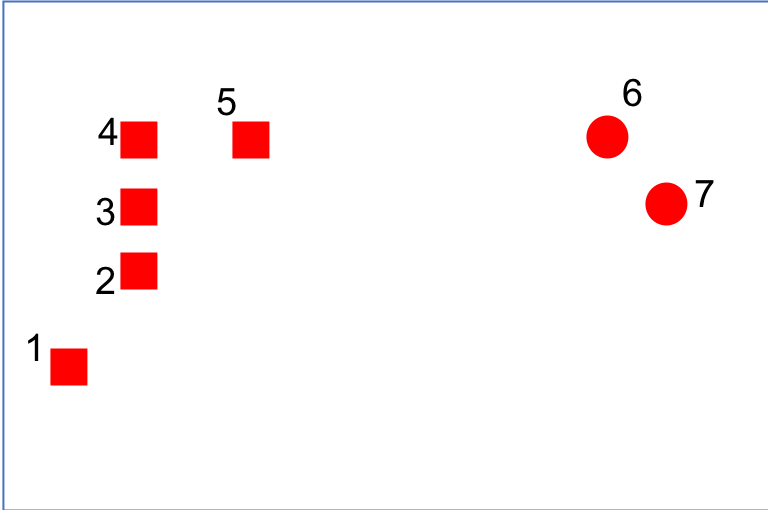


Figure 1 example - initialization

Next assume that points 1, 6, and 7 as initial cluster centers.

At the end of first iteration points 1,2,3 and 4 will be in one cluster. 5 and 6 will be in another cluster. And 7 will be in the last cluster. Note here that the distance between 1 and 5 is larger than the distance between 5 and 6 and so 5 is assigned to the cluster represented by 6 (blue cross). Update the cluster centers and the following figure 2 shows the centers (approximately) and the clusters at the end of first step.

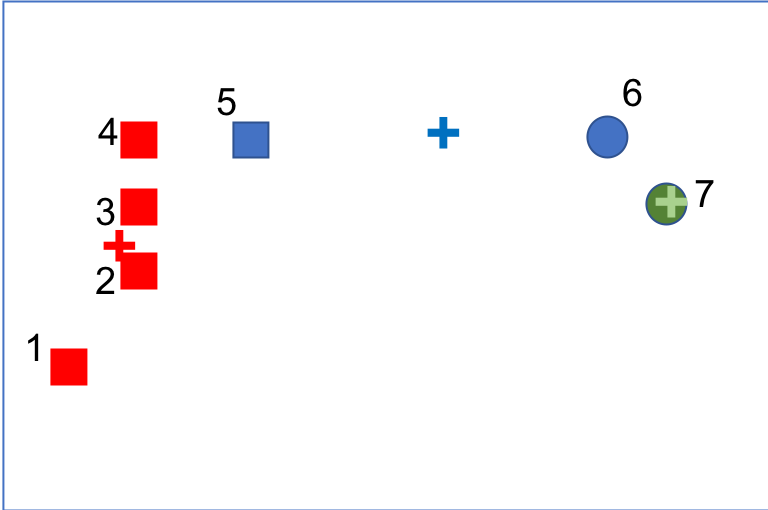


Figure 2 example - after one iteration

Now, the cluster center for the red cluster moved closer to point 5 due to 2, 3 and 4. And the cluster center for the blue cluster moved away from 6 due to point 5. In the next iteration point 5 will decide that it is closer to the red cluster and point 6 will decide that it is closer to the green cluster. This will cause blue cluster to be empty as shown in figure 3.

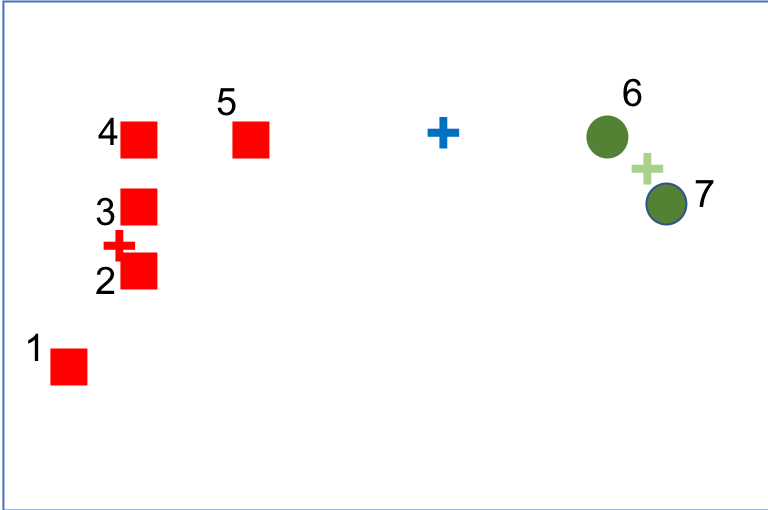


Figure 3 example - after 2nd iteration

### It is impossible to have non-convex clusters. In figure 4, and are cluster centers of two clusters, and are two clusters that share a common border, and the black dashed line is the bisector of the two clusters. We choose 2 data points and from randomly. They must lie on the same side of the bisector with since they are closer to than . If we take any between and , by geometry it must lie on the same side of the bisector with and , and thus it must be assigned to , i.e. the cluster is convex.

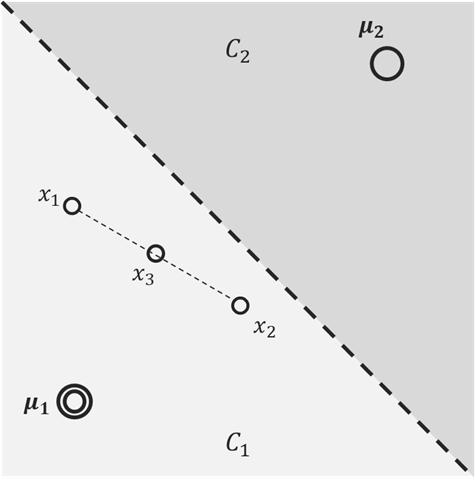


Figure 4

## (35 points) Expectation-Maximization (EM) for Mixture Model Clustering

Where

Log likelihood

|  |  |
| --- | --- |
| Algorithm 3.1 General Mixture Model | |
|  | Initialize  **Repeat**  **for all** **do**    **end for**  **for** **do**      **end for**    **if**    **else**  exit |

### Replace in with that of Bernoulli distribution:

To find the MLE estimator, let , i.e.: