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 to it, will decrease if changes or will not change if does not change. Therefore, the objective function will always decrease or not change.

(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.

### The is possible. A case is when the cluster number user assigned is even greater than the number of data points. …

### It is impossible to have non-convex clusters. In figure 2.1, and are cluster centers of two clusters, and , that share a common border, and the grey dashed line is the bisector of the two clusters. Then we choose 2 data points and from randomly, and 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.

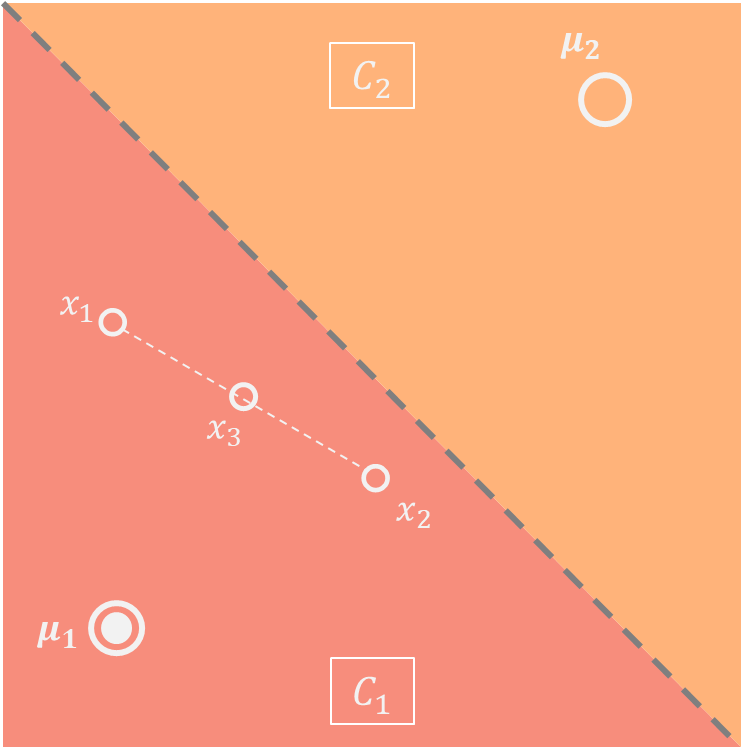


Figure 2.1

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

|  |  |
| --- | --- |
| 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.: