Some Algorithms for Unsupervised Clustering

Jussi Tohka
Outline

- Unsupervised classification problem
- The K-means algorithm
- The EM-algorithm
Unsupervised classification problem

Classification procedures that use labeled samples to train the classifier are said to be *supervised*.

Sometimes we do not have the training data. Classification procedures which use only unlabeled samples are said to be *unsupervised*.

An unsupervised procedure must use the unlabeled input data to estimate the parameter values for the classification problem at the hand and also to classify the data.
The K-means algorithm

- An algorithm to estimate the unknown cluster centres (means) $M = \{\mu_1, \ldots, \mu_K\}$ based on the data $D = \{x_1, \ldots, x_N\}$.

- Aims to minimize

$$J(M) = \sum_{i=1}^{N} ||x_i - \mu_i||^2,$$

where $\mu_i$ is the closest cluster centre to $x_i$. 

Some Algorithms for Unsupervised Clustering – p.4/30
The K-means algorithm

1. Initialize $K, \mu_j$
2. Classify all the data $x_i, i = 1, \ldots, N$ according to the nearest $\mu_j$.
3. Recompute each $\mu_j$.
4. If no change then terminate, otherwise go to step 2.
Properties of the K-means algorithm

- Simple
- A local optimization algorithm
- The cost function $J(M)$ is not be suitable for many problems.
  e.g. for separating clusters of different size.
The K-means algorithm: An experiment
The K-means algorithm: An experiment
The K-means algorithm: An experiment
K-means experiment, iteration 1
K-means experiment, iteration 1
K-means experiment, iteration 1
K-means experiment, iteration 1
K-means experiment, iteration 2
K-means experiment, iteration 2
K-means experiment, iteration 2
K-means experiment, iteration 3
K-means experiment, iteration 3
K-means experiment, iteration 3
K-means experiment, iteration 4
K-means experiment, iteration 4

Some Algorithms for Unsupervised Clustering – p.11/30
K-means experiment, iteration 4
K-means experiment, iteration 5
K-means experiment, iteration 6
K-means experiment, iteration 6
K-means experiment, iteration 6
K-means experiment, iteration 7
K-means experiment, iteration 7
K-means experiment, iteration 8
K-means experiment, iteration 8
K-means experiment, iteration 8
K-means experiment, iteration 9
K-means experiment, iteration 9
K-means experiment, iteration 9
K-means experiment, iteration 10
K-means experiment, iteration 10

Some Algorithms for Unsupervised Clustering – p.17/30
K-means experiment, iteration 10
K-means experiment, iteration 11

Labels did not change Convergence!

Some Algorithms for Unsupervised Clustering – p.18/30
K-means experiment, iteration 11
K-means experiment, iteration 11

Labels did not change $\rightarrow$ Convergence!
K-means experiment, results

True labels

Clustering result
K-means experiment, results

True labels

Errors, Error rate 1 %
K-means cost function values

![Graph showing K-means cost function values](image-url)
K-means cost function values

![Graph showing K-means cost function values](image-url)
Examples where K-means doesn’t work

True labels
Examples where K-means doesn’t work

True labels

K-means result
Examples where K-means doesn’t work

True labels
Examples where K-means doesn’t work
EM-algorithm for clustering

- Aims to maximize the likelihood of the mixture density given the data, i.e.

\[ p(D|\theta) = \prod_{i=1}^{N} \sum_{j=1}^{K} P_j \cdot g_j(x_i|\mu_j, \Sigma_j), \]

where \( g_j(\cdot) \) is the normal density with the mean \( \mu_j, \Sigma_j \), and \( P_j \) is the prior probability of the class \( j \).

- Maximization of the above likelihood can be done with an EM-algorithm. This algorithm is also described in the book but it is not called an EM-algorithm there.
EM-algorithm for clustering

1. Initialize $K, \mu_j^0, \Sigma_j^0, P_j^0$.

2. (E-step) Compute the probabilities $p_{ij}^{t+1}$ of $x_i$ belonging to the class $j$ based on the parameter values $\mu_j^t, \Sigma_j^t, P_j^t$ from the previous iteration.

3. (M-step) Re-compute the parameter values:

$$P_j^{t+1} = \frac{1}{n} \sum_i p_{ij}^{t+1}; \quad \mu_j^{t+1} = \frac{\sum_i p_{ij}^{t+1} x_i}{\sum_i p_{ij}^{t+1}}$$

$$\Sigma_j^{t+1} = \frac{\sum_i p_{ij}^{t+1} (x_i - \mu_j)(x_i - \mu_j)^T}{\sum_i p_{ij}^{t+1}}$$

4. Terminate if converged, else go to step 2.
EM versus K-means
EM versus K-means

True labels

K-means result
EM versus K-means

True labels

EM result

Some Algorithms for Unsupervised Clustering – p.25/30
EM versus K-means

True labels
EM versus K-means

True labels

K-means result
EM versus K-means

True labels

EM result
EM versus K-means

True labels
EM versus K-means

True labels

K-means result

Some Algorithms for Unsupervised Clustering – p.27/30
EM versus K-means

True labels

EM result
Application: MRI-segmentation
Application: MRI-segmentation
Application: MRI-segmentation

Ground truth

EM,T1+T2

K-means,T1+T2
Application: MRI-segmentation

Ground truth  EM, T1+T2  EM, only T1
Application: MRI-segmentation