

Gaussian Mixture Models

1 Questions

1. Which quantities define a multi-variate Gaussian distribution?
2. What information about multi-dimensional data can be seen in Σ ?
3. What is the advantage of adding variance information to a codebook vector?
4. Which quantities define a GMM?
5. What is the cost function of the EM algorithm?
6. What are the two main steps of the EM algorithm in each iteration?
7. How do these steps compare against the two iteration steps in the k-means algorithm? What does the introduced latent variable have to do with this difference?
8. How can a GMM / the EM algorithm be initialized?
9. What are abort conditions of the EM algorithm?
10. What is a pitfall of the EM algorithm and how can it be circumvented?

2 Answers

1. The mean vector μ and the covariance matrix Σ define a Gaussian distribution.
2. The diagonal of the covariance matrix tells us how "far" the data points are spread out into each dimension. The other elements give us information about the data correlation between the different dimensions.
3. The variance information allows us to move the decision line / plane according to the actual data distribution. When attempting classification, this can result in better performance.
4. Mean vector μ_k and the covariance matrix Σ_k for each Gaussian. Additionally, the weights g_k for each Gaussian are required.
5. The observation probability of the data given the GMM is to be maximized.
6. Each iteration consists of an Expectation and a Maximization step.
7. The expectation step corresponds to the classification and the maximization step corresponds to correction in the k-means algorithm (but of course the actual operations are different). However since we are working with probabilities, the classification does not perform a hard decision but a soft decision in form of probabilities. The introduced latent variable helps with this process.
8. The GMM can be initialized applying a codebook training (k-means) to the data set. The codebook vectors give the initial means of the Gaussians. A codebook classification (hard decision) is then used to select which data points are used for each Gaussian to estimate the covariance and the weight.
9. Maximum number of iterations. Threshold for relative cost function improvement.
10. If only few data points are available for training, it can happen that the variance of the Gaussian approaches zero. Although this result is good according to the cost function, the data points are generally not well represented by this solution. Consequently a lower bound for the allowed variances can be introduced.