4.5 GAUSSIAN MIXTURE MODELS AND EXPECTATION MAXIMIZATION

Gaussian Mixture Models is a "soft" clustering algorithm, where each point probabilistically "belongs" to all clusters. This is different than k-means where each point belongs to one cluster.

- The Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mix of Gaussian distributions with unknown parameters.
- For example, in modeling human height data, height is typically modeled as a normal distribution for each gender with a mean of approximately 5'10" for males: and 55" for females. Given only the height data and not the gender assignments for each data point, the distribution of all heights would follow the sum of two scaled (different variance) and shifted (different mean) normal distributions
- A model making this assumption is an example of a Gaussian mixture model.
- Gaussian mixture models do not rigidly classify each and every instance into one class or the other. The algorithm attempts to produce K-Gaussian distributions that would take into account the entire training space.

Every point can be associated with one or more distributions. Consequently, the deterministic factor would be the probability that each point belongs to a certain Gaussian distribution.

GMMs have a variety of real-world applications. Some of them are listed below.

a) Used for signal processing

b) Used for customer churn analysis

c) Used for language identification

d) Used in video game industry

e) Genre classification of songs

4.5.1 Expectation-maximization

In Gaussian mixture models, an expectation-maximization method is a powerful tool for estimating the parameters of a Gaussian mixture model. The expectation is termed E and maximization is termed M.

- Expectation is used to find the Gaussian parameters which are used to represent each component of gaussian mixture models. Maximization is termed M and it is involved in determining whether new data points can be added or not.
- The Expectation-Maximization (EM) algorithm is used in maximum likelihood estimation where the problem involves two sets of random variables of which one, X, is observable and the other, Z, is hidden.
- The goal of the algorithm is to find the parameter vector Φ that maximizes the likelihood of the observed values of X, L(Φ/ X).
- But in cases where this is not feasible, we associate the extra hidden variables Z and express the underlying model using both, to maximize the likelihood of the joint distribution of X and Z, the complete likelihood $L_C(\Phi / X,Z)$.
- Expectation-maximization (EM) is an iterative method used to find maximum likelihood estimates of parameters in probabilistic models, where the model depends on unobserved, also called latent, variables.
- EM alternates between performing an expectation (E) step, which computes an expectation of the likelihood by including the latent variables as if they were observed, and maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found in the E step.
- The parameters found on the M step are then used to start another E step, and the process is repeated until some criterion is satisfied. EM is frequently used for data clustering like for example in Gaussian mixtures.
- In the **Expectation step**, find the expected values of the latent variables (here you need to use the current parameter values)
- In the **Maximization step**, first plug in the expected values of the latent variables in the log-likelihood of the augmented data. Then maximize this log-likelihood to reevaluate the parameters
- Expectation-Maximization (EM) is a technique used in point estimation. Given a set of observable variables X and unknown (latent) variables Z we want to estimate parameters ɵ in a model.

- The expectation maximization (EM) algorithm is a widely used maximum likeli-hood estimation procedure for statistical models when the values of some of the variables in the model are not observed
- The EM algorithm is an elegant and powerful method for finding the maximum likelihood of models with hidden variables. The key concept in the EM algorithm is that it iterates between the expectation step (E-step) and maximization step (M-step) until convergence.
- In the E-step, the algorithm estimates the posterior distribution of the hidden variables Q given the observed data and the current parameter settings, and in the M-step the algorithm calculates the ML parameter settings with Q fixed.
- At the end of each iteration the lower bound on the likelihood is optimized for the given parameter setting (M-step) and the likelihood is set to that bound (E-step), which guarantees an increase in the likelihood and convergence to a local maximum, or global maximum if the likelihood function is unimodal.
- Generally, EM works best when the fraction of missing information is small and the dimensionality of the data is not too large. EM can require many iterations, and higher dimensionality can dramatically slow down the E-step.
- EM is useful for several reasons: conceptual simplicity, ease of implementation, and the fact that each iteration improves $l(\theta)$. The rate of convergence on the first few steps is typically quite good, but can become excruciatingly slow as you approach local optima.
- Sometimes the M-step is a constrained maximization, which means that there are constraints on valid solutions not encoded in the function itself.
- Expectation maximization is an effective technique that is often used in data analysis to manage missing data. Indeed, expectation maximization overcomes some of the limitations of other techniques, such as mean substitution or regression substitution. These alternative techniques generate biased estimates and, specifically, underestimate the standard errors. Expectation maximization overcomes this problem.