Expectation Maximisation

The EM algorithm is a widely used approach for learning in the presence of unobserved variables, provided the general probability distribution governing these variables is known.

Estimating Means of k Gaussians

The problem basically is that there is a set of data D that originates from k distinct Gaussian distributions. Each instance is generated by first selecting a Gaussian at random and then a single instance x is selected from that distribution. The variance of the Gaussians is known, and the task is to output a hypothesis that describes the means of each of the k distributions.

For a single distribution, this mean is given by

$$mean_{ML} = \underset{mean}{argmin} \sum_{i=1}^{m} \left(x_i - \mu \right)^2$$

But since the data comes from a mixture of Gaussians this is not valid.

We define each instance as a triplet <x,z1,z2..zN> where x is the observed value of an instance and z1 indicates the normal distribution that was used to generate the value x.

If zi has the value 1, the value x was generated from the ith Gaussian.

The EM algorithm searches for a maximum likelihood hypothesis by repeatedly re-estimating the values of the hidden variables z given its current hypothesis, then recalculating the maximum likelihood hypothesis using these expected values for the hidden variables.

The algorithm is given as

1. Calculate the expected value E[zi] of each hidden variable z, assuming the current hypothesis <mean1, mean2…> holds.
2. Calculate a new maximum likelihood hypothesis h’ assuming the value taken by each variable z is the expected value.
3. Replace h by h’ and re-iterate.

The expected value E[zi] is calculated as

$$E[z_{ij}] = \frac{p\left(x = x_i | \mu = \mu_j\right)}{\sum_{n=1}^{k} p\left(x = x_i | \mu = \mu_n\right)}$$

$$E[z_{ij}] = \frac{e^{-\frac{1}{2\sigma^2}\left(x - \mu_i\right)^2}}{\sum_{i=1}^{k} e^{-\frac{1}{2\sigma^2}\left(x_i - \mu_n\right)^2}}$$

The maximum likelihood hypothesis is a collection of all the means calculated using this expected value, where each mean is given by

$$\mu_j = \frac{\sum_{i=1}^{m} E[z_{ij}] x_i}{\sum_{i=1}^{m} E[z_{ij}]}$$

The General Statement of the EM algorithm

1. Estimation step E

$$Q(h' | h) = E\left[\ln P(Y | h') | h, X\right]$$

2. Maximization step M

$$h = \underset{h'}{argmax}\ Q(h' | h)$$

Where

θ = Set of parameters = <mean1,mean2...>

θ' = Set of revised parameters

X = observed data in a set of m independently drawn instances

Z = unobserved data in these same instances

Y = random variable defined in terms of Z

h = Initial hypothesis

h' = Revised hypothesis

When the function Q is continuous, the EM algorithm converges to a stationary point of the likelihood function P. When this likelihood function has a single maximum, EM converges to this global maximum likelihood for h’. Otherwise, it guarantees only a local maximum.