Title: EM algorithm and GMM model

URL: https://en.wikipedia.org/wiki/EM_algorithm_and_GMM_model

PageID: 64563432

Categories: Category:Machine learning, Category:Regression models

---

In statistics, EM (expectation maximization) algorithm handles latent variables, while GMM is the Gaussian mixture model.

Background

In the picture below, are shown the red blood cell hemoglobin concentration and the red blood cell volume data of two groups of people, the Anemia group and the Control Group (i.e. the group of people without Anemia). As expected, people with Anemia have lower red blood cell volume and lower red blood cell hemoglobin concentration than those without Anemia.

x {\displaystyle x} is a random vector such as x := (red blood cell volume , red blood cell hemoglobin concentration) {\displaystyle x:={\big (}{\text{red blood cell volume}},{\text{red blood cell hemoglobin concentration}}{\big)}} , and from medical studies [citation needed] it is known that x {\displaystyle x} are normally distributed in each group, i.e. x ~ N (μ , Σ) {\displaystyle x\sim {\mathcal {N}}(\mu ,\Sigma)} .

z {\displaystyle z} is denoted as the group where x {\displaystyle x} belongs, with z i = 0 {\displaystyle z_{i}=0} when x i {\displaystyle x_{i}} belongs to Anemia Group and z i = 1 {\displaystyle z_{i}=1} when x i {\displaystyle x_{i}} belongs to Control Group. Also z ~ Categorical ■ (k , φ) {\displaystyle z\sim \operatorname {Categorical} (k,\phi)} where k = 2 {\displaystyle k=2} , φ j ≥ 0 , {\displaystyle \phi _{j}\geq 0,} and ∑ j = 1 k φ j = 1 {\displaystyle \sum _{j=1}^{k}\phi _{j}=1} . See Categorical distribution .

The following procedure can be used to estimate φ , μ , Σ {\displaystyle \phi ,\mu ,\Sigma } .

A maximum likelihood estimation can be applied:

As the z i {\displaystyle z_{i}} for each x i {\displaystyle x_{i}} are known, the log likelihood function can be simplified as below:

Now the likelihood function can be maximized by making partial derivative over μ , Σ , φ {\displaystyle \mu ,\Sigma ,\phi } , obtaining:

If z i {\displaystyle z_{i}} is known, the estimation of the parameters results to be quite simple with maximum likelihood estimation . But if z i {\displaystyle z_{i}} is unknown it is much more complicated.

Being z {\displaystyle z} a latent variable (i.e. not observed), with unlabeled scenario, the Expectation Maximization Algorithm is needed to estimate z {\displaystyle z} as well as other parameters. Generally, this problem is set as a GMM since the data in each group is normally distributed. [circular reference]

In machine learning , the latent variable z {\displaystyle z} is considered as a latent pattern lying under the data, which the observer is not able to see very directly. x i {\displaystyle x_{i}} is the known data, while φ , μ , Σ {\displaystyle \phi ,\mu ,\Sigma } are the parameter of the model. With the EM algorithm, some underlying pattern z {\displaystyle z} in the data x i {\displaystyle x_{i}} can be found, along with the estimation of the parameters. The wide application of this circumstance in machine learning is what makes EM algorithm so important.

EM algorithm in GMM

The EM algorithm consists of two steps: the E-step and the M-step. Firstly, the model parameters and the z (i) {\displaystyle z^{(i)}} can be randomly initialized. In the E-step, the algorithm tries to

guess the value of z (i) {\displaystyle z^{(i)}} based on the parameters, while in the M-step, the algorithm updates the value of the model parameters based on the guess of z (i) {\displaystyle z^{(i)}} of the E-step. These two steps are repeated until convergence is reached.

The algorithm in GMM is:

Repeat until convergence:

With Bayes Rule , the following result is obtained by the E-step:

p (z (i) = j | x (i) ; φ , μ , Σ) = p (x (i) | z (i) = j ; μ , Σ) p (z (i) = j ; φ) ∑ l = 1 k p (x (i) | z (i) = l ; μ , Σ) p (z (i) = l ; φ) {\displaystyle p\left(z^{(i)}=j|x^{(i)};\phi ,\mu ,\Sigma \right)={\frac {p\left(x^{(i)}|z^{(i)}=j;\mu ,\Sigma \right)p\left(z^{(i)}=j;\phi \right)}{\sum _{l=1}^{k}p\left(x^{(i)}|z^{(i)}=l;\mu ,\Sigma \right)p\left(z^{(i)}=l;\phi \right)}}}

According to GMM setting, these following formulas are obtained:

p (x (i) | z (i) = j ; μ , Σ) = 1 (2 π) n / 2 | Σ j | 1 / 2 exp ■ (– 1 2 (x (i) – μ j) T Σ j – 1 (x (i) – μ j)) {\displaystyle p\left(x^{(i)}|z^{(i)}=j;\mu ,\Sigma \right)={\frac {1}{(2\pi)^{n/2}\left|\Sigma _{j}\right|^{1/2}}}\exp \left(-{\frac {1}{2}}\left(x^{(i)}-\mu _{j}\right)^{T}\Sigma _{j}^{-1}\left(x^{(i)}-\mu _{j}\right)\right)}

p (z (i) = j ; φ) = φ j {\displaystyle p\left(z^{(i)}=j;\phi \right)=\phi _{j}}

In this way, a switch between the E-step and the M-step is possible, according to the randomly initialized parameters.

References