**Expectation Maximization Algorithm with Combinatorial Assumption**

**Abstract**

**Keywords:**

**1. Introduction**

It is not easy to specified the mapping *φ*(*X*) in traditional EM or the joint PDF *f*(*X*, *Y*) in practical EM. Therefore, the purpose of this research is to extend competency of EM, in which the observed data *Y* is assumed to be a combination of *X*, called combinatorial assumption (CA). In other words, *Y* can be calculated by an analytic function of *X*, which is more feasible than specifying the mapping *φ*(*X*) and easier than specifying the joint PDF *f*(*X*, *Y*). The analytic function is arbitrary but it is linear called regression function in this research for convenience and feasibility. In many cases, it is possible to convert some analytic functions into linear functions. The next section is the main one which focuses on EM with CA. In general, this research is an extension of EM algorithm.

**2. EM with combinatorial assumption**

Without loss of generality, suppose every random vector variable *Yi* degrades into random scalar variable *yi* and thus, the condition expectation *Q*(Θ | Θ’) becomes:

Of course, we had the sample = {*y*1, *y*2,…, *yN*} of size *N* in which all *yi* (s) are mutually independent and identically distributed (iid). Suppose *X* = (*x*1, *x*2,…, *xn*) which is vector of size *n* distributes normally with mean vector *μ* and covariance matrix Σ as follows:

Where the superscript “*T*” denotes vector (matrix) transposition operator. Let *y* be random variable representing all sample random variable *yi*. Suppose there is an assumption that *y* is a combination of partial random variables (components) of as follows:

This implies the equation above is regression function in which *y* is called responsor and each *xi* is called regressor whereas *αj* are called regressive coefficients. The assumption is combinatorial assumption (CA) aforementioned. As a convention, let

The regression function is re-written as follows:

Suppose *y* distributed normally with mean and variance *σ*2 as follows:

The density probability function (PDF) of *y* is now defined by support of regression model as follows:

The equation above is not really total probability rule but it implies that the conditional PDF *f*(*y* | *X*) is substituted by the regression model. Consequently, the condition expectation *Q*(Θ | Θ’) is becomes:

Where parameter Θ = (*μ*, Σ, *α*, *σ*2)*T*. It is necessary to specify the conditional PDF *f*(*X* | *y*, Θ). Indeed, we have:

Let *g*(*X*, *y* | *μ*, Σ, *α*, *σ*2) be the numerator of *f*(*X* | *y*, Θ):

Where,

The expression is approximated with *μ* as follows:

As a result, *g*0(*X*, *y* | *μ*, Σ, *α*, *σ*2) and *g*(*X*, *y* | *μ*, Σ, *α*, *σ*2) is approximated by:

Of course, we have:

Let *A* be the denominator of *f*(*X* | *y*, Θ) which is the integral of *g*(*X*, *y* | *μ*, Σ, *α*, *σ*2) over *X*:

Where *B* is the integral of *g*0(*X*, *y* | *μ*, Σ, *α*, *σ*2) over *X*:

It requires to calculate *B* to determine *f*(*X* | *y*, Θ). Due to shifted Gaussian integral:

We have:

Thus, *A* is approximated as follows:

As a result, the PDF *f*(*X* | *y*, Θ) is approximated as follows:

Let *k*(*y*|Θ) be the constant with subject to *X*, which is defined as follows:

Shortly, the conditional PDF *f*(*X* | *y*, Θ) is specified (approximated) at E-step of some *t*th iteration process as follows:

Consequently, the condition expectation *Q*(Θ | Θ(*t*)) at some *t*th iteration is totally determined:

At M-step of the current *t*th iteration, *Q*(Θ|Θ(*t*)) is maximized by setting its partial derivatives regarding Θ to be zero. The first-order partial derivative of *Q*(Θ | Θ(*t*)) with regard to *μ* with suppose that *Q*(Θ | Θ(*t*)) is analytic function is:

Due to shifted Gaussian integral:

We have:

Note, Σ is invertible and symmetric. The next parameter *μ*(*t*+1) at E-step of some *t*th iteration that maximizes *Q*(Θ|Θ(*t*)) is solution of the equation as follows:

Where,

And

The first-order partial derivative of *Q*(Θ | Θ(*t*)) with regard to Σ with suppose that *Q*(Θ | Θ(*t*)) is analytic function is:

The next parameter Σ(*t*+1) that maximizes *Q*(Θ|Θ(*t*)) is the solution of equation formed by setting to zero matrix. Let (**0**) denote zero matrix.

We have:

(Because *μ* is replaced by *μ*(*t*))

Therefore, the next parameter Σ(*t*+1) is obtained:

Due to shifted Gaussian integral:

We have:

The first-order partial derivative of *Q*(Θ | Θ(*t*)) with regard to *α*0 with suppose that *Q*(Θ | Θ(*t*)) is analytic function is:

Due to shifted Gaussian integral:

We obtain:

Therefore, the next parameter *α*0(*t*+1) is obtained by setting the partial derivative to be zero:

Note, is replaced by at current *t*th iteration. The first-order partial derivative of *Q*(Θ | Θ(*t*)) with regard to with suppose that *Q*(Θ | Θ(*t*)) is analytic function is:

Replacing *μ* and *α*0 by *μ*(*t*) and *α*0(*t*), respectively at current *t*th iteration along with applying shifted Gaussian integral, we have:

Therefore, the next parameter is obtained by setting the partial derivative to be zero:

Note, the superscript “–1” denotes matrix inversion. The first-order partial derivative of *Q*(Θ | Θ(*t*)) with regard to *σ*2 with suppose that *Q*(Θ | Θ(*t*)) is analytic function is:

Approximating *α*0, , and *X* by next parameters *α*0(*t*+1), , and *μ*(*t*+1), respectively we have:

Therefore, the next parameter is obtained by setting the partial derivative to be zero:

**3. Numerical simulation**

**4. Discussions and conclusions**

**References**