

# Mixture Regression Model for Incomplete Data

N. P. Lộc<sup>1</sup>

**Abstract:** The Regression Expectation Maximization (REM) algorithm, which is a variant of Expectation Maximization (EM) algorithm, uses parallelly a long regression model and many short regression models to solve the problem of incomplete data. Long regression model is entire regression function which is the resulted model and short regression models are partial regression functions which are inverses of entire regression function. I proposed REM in a different research in which one entire regression function is built parallelly with many partial inverse regression functions and then missing values are fulfilled by expectations relevant to both entire regression function and inverse regression functions. Experimental results proved resistance of REM to incomplete data, but accuracy of REM decreases insignificantly when data sample is made sparse with loss ratios up to 80%. Like traditional regression analysis methods, accuracy of REM can be decreased if data varies complicatedly with many trends. In this research, I propose a so-called Mixture Regression Expectation Maximization (MREM) algorithm. MREM is full combination of REM and mixture model in which I use two EM processes in the same loop. MREM uses the first EM process for exponential family of probability distributions to estimate missing values as REM does. Consequently, MREM uses the second EM process to estimate parameters as mixture model method does. The purpose of MREM is to take advantages of both REM and mixture model. Unexpectedly, experimental result shows that MREM is less accurate than REM. I try to weight partial models of MREM by product of component probabilities and conditional probabilities or to select most appropriate partial model in order to improve estimation accuracy, but the final results are not as good as expected. However, MREM is essential because a different approach for mixture model can be referred by fusing linear equations of MREM into a unique curve equation proposed by some other researches.

---

<sup>1</sup> Independent scholar  
1/4B Ton Duc Thang, Long Xuyen, An Giang 881092, Vietnam  
*ng\_phloc@yahoo.com*