Available online at www.joac.info

ISSN: 2278-1862

Journal of Applicable Chemistry

2016, 5(5): 908-1033

(International Peer Reviewed Journal)

Research tutorial (ResT)

[Computational/Chemical]TensorLab(CTLab)
Part 2: Linear Least squares in Matlab

K. Ramakrishna[1] and R. Sambasiva Rao[2*]

1. Department of Chemistry, Gitam Institute of Science, Gitam University, Visakhapatnam, 530 017, **INDIA**
2. School of Chemistry, Andhra University, Visakhapatnam 530 003, **INDIA**

Email: karipeddirk@gmail.com, rsr.chem@gmail.com

Accepted on 10th September2016

(Dedicated with profound respects to Dr K V Suryanarayana, former professor of statistics, Andhra University on his seventy fifth birth anniversary)

Conspectus

Background:The models are precise expression of experimental results but not at all a substitute.On the other hand, data driven models do not start with any prefixed model, but at the end a model emerges.The linguistic models or automatic genetic algorithm/genetic programming generate a set of equivalent models and submerge most of earlier category although the mathematical/physical form is different for a naked eye. The regression models, self-organizing models, multiple-(constrained) optimization (with conflicting subgoal) models form major category in the bandwagon of computational tools.

Purpose: The focus of the current research review is to start with simple as possible matrix formulae to estimate regression parameters of linear/polynomial models in one explanatory variable (x) and coding in Matlab illustrating the application for small number of (six to ten) noise free simulated data. The results can be arrived at without any gadget. The perturbation of statistics of model parameters with (homoscedastic) Gaussian noise is dilated. The effects of outliers are exemplified remedial measures viz. least median squares (LMS) and least trimmed squares (LTS) are illustrated. The exhaustive set of models in analyzing data from polynomial models is developed in polyLS2015. The method of least squares is derived for univariate replicate data adhering to mean model perturbed by Gaussian noise. MAD statistic, a robust measure of central tendency is used to detect outliers and probe into central tendency of data in their presence. Linear parametricRegression with Multiple-X variables (MLR) and single response, a hard model is considered. A function of two explanatory variables is coded in MLR2015.m and simulated data sets amply illustrate its utility.In this phase, only mathematical formulae, m-functions, simple-as-possible examples are narrated. Anobject with typical results of each method is invoked and tabular and graphic output programs are available.

In the second phase, the default datasets, autotest_$$$ for all possible testing of program capabilities are discussed. The knowledge-based approach for input checking, validating input data/intermediate results structure to process a mathematical task (set of formulae) are developed in the if-then-else numericalrules. The necessary conditions, failure flags, remedial measures for each type of analysis and textual summary of algorithm flow with Matlab functions used are narrated in the next phase.

Each of these modules run individually and a bunch of them form a combined solution.A template (GUI) mode for selection and transparent flow of the software is under development.

Keywords: Cause-effect relationships, normal distribution, homoscedastic, heteroscedastic, residuals, absolute_residuals, squares of residuals, least_sum, regression parameters, linear, normal population, LLS, LAD, polyLS, LMS, MLR, statistics [residuals, parameters], ANOVA, information, KBs, NCs, failure conditions, Remedial measures, matlab_functions.
