Codes for BigDat19 Lectures

Soumya D. Mohanty

Dept. of Physics and Astronomy

The University of Texas Rio Grande Valley

Abstract

This is a user manual for codes associated with the course "Swarm intelligence methods for statistical regression" delivered at the BigDat 2019 winter school on big data, Cambridge, UK. The codes are written in the Matlab programming language and have been tested in Matlab R2018b.

I. INTRODUCTION

In the lectures delivered under the course "Swarm intelligence methods for statistical regression" at the BigDat 2019 international winter school on Big Data, Cambridge University, UK, we consider some workhorse statistical regression problems for illustrating the application of Particle Swarm Optimization (PSO). This document provides a user manual for and brief descriptions of these codes.

Please report bugs to: `soumya.mohanty@utrgv.edu`

II. CONVENTIONS

Most of the code consists of Matlab [1] functions with names that follow the pattern "crcb⟨name⟩.m". (The prefix stands for CRC book and refers to the book [2] on which the course is based.)

Function names not preceded by this prefix are actually borrowed from other code libraries developed by the author. These have been ported here to create a standalone package and are named differently for the purpose of code maintenance. The user should not need to look inside such functions in the normal course of using the top level codes.

Some Matlab scripts have been provided, with file names that start with "test_crcb⟨name⟩", that reproduce some of the main results shown in the lectures. These scripts can serve as templates that readers can use to write their own codes for calling the functions provided.

Each function has an associated help that can be invoked by typing "help <function name>" at the Matlab command prompt. The help style emulates the one followed by native Matlab functions and provides usage instructions and, in some cases, examples of usage.

In addition to the help, each function is documented internally, although no claims are made as to the quality of said documentation. However, we hope that in most cases, the documentation will be adequate for understanding the code should the reader feel the need to make modifications.

III. LIST OF FUNCTIONS

We first list the functions starting with the "crcb" prefix (this prefix is ommitted from the name of the function) followed by the non-crcb functions. We then provide usage instructions for the "test_crcb⟨name⟩" scripts. Where appropriate, the description of a function points the reader to the lectures [3–5] where a discussion of the relevant topic can be found. The lecture slides are available in the Google drive folder shared with the attendees of the conference.

CRCB function	**Description**
pso	Implements local best (*lbest*) PSO[5]. It is assumed that the search space is a hypercube with the coordinate along each dimension lying in the range $[0, 1]$ (standardized coordinates). The function provides flexibility in changing the parameters associated with the *lbest* PSO algorithm and it can be used on any fitness function provided the latter has a specified interface.
psotestfunc	Implements the generalized Rastrigin fitness function. The primary purpose of this function is to serve as a template for creating new fitness functions that can be fed to the pso function.
genqcsig	Generates the quadratic chirp signal used in the parametric regression problem [3, 5]. The amplitude and phase parameters of the signal can be specified through the input arguments to this function
genbsplsig	Generates the single B-spline signal used in the non-parametric regression problem [3, 5]. The breakpoints for the B-spline can be specified through the input arguments to this function.
genqcdata	Generate a single data realization containing a quadratic chirp signal added to a white, Gaussian noise realization.
genbspldata	Generate a single data realization containing a single B-spline signal added to a white, Gaussian noise realization.
qcfitfunc	The fitness function for the quadratic chirp regression problem. Returns the sum of squared residuals after maximizing [5] over the amplitude parameter (A). See [2] for the derivation of this fitness function.

regsplfitfunc	The fitness function for the regression spline problem (without the penalized spline regulator). Returns the sum of squared residuals after minimizing it over B-spline coefficients [5].
qcpso	Applies the pso function using the Best-of-M-runs strategy [4] to the quadratic chirp regression fitness function (qcfitunc). It is assumed that the Matlab Parallel Computing Toolbox is available for implementing the BMR strategy. If not, the "parfor" loop should be changed to a "for" loop.
regsplpso	Applies the pso function using the BMR strategy [4] to the regression spline fitness function (regsplfitfunc). It is assumed that the Matlab Parallel Computing Toolbox is available for implementing the BMR strategy. If not, the "parfor" loop should be changed to a "for" loop.
crdnlsplfit	Implements the fitness function for a cardinal spline fit. Note that this fitness function is not to be optimized by pso. It returns the best fit curve.
chkstdsrchrng	Helper function that checks if a given location (standardized coordinates) is inside the unit hypercube. Returns a logical output indicating failure or success of this condition.

Test function	Description
pso	Shows how to call the pso function using the psotestfuncfitness function as the target fitness function to be optimized.
qcpso	Shows how to call pso on qcfitfunc.
regsplpso	Shows how to call pso on regsplfitfunc.
crdnlsplfit	Shows how to call the crdnlslfit function.

[1] Matlab R2018b, Mathworks. URL: https://www.mathworks.com.

[2] Soumya D. Mohanty, *Swarm intelligence methods for statistical regression* (Chapman and Hall/ CRC press, 2018).

[3] Lecture 1 slides.

[4] Lecture 2 slides

[5] Lecture 3 slides