Hand-on Exercise 3

Guidelines

(1) Each group consists of at most two students. Members in the same team will receive the same grade. (Please give your group member names to TA.)

(2) The deadline for turning in your report is April 2 (Thursday) 1:00 PM.

Assignment

(1) Each team's task is to unveil the embedded information of three given datasets.

(2) **Three datasets** are posted on the course page.

(3) Each team may use MATLAB or Python as the development platform.

(4) The performance will be judged by the quality of solutions obtained, the associated analysis and discussions.

(5) For dataset "data1-ex3-s25.csv", perform the following tasks:

(i) (5 point) Use the least-squares-estimation linear regression method (page 15 of Lecture 4) to find the hidden linear relationship $R1$ between the input and output variables. Record the computational time and compute the value of MSE (mean squared errors).

(ii) (5 points) Use the LSSVR model (page 64 of Lecture 5) with parameters $C = 1$ and $\epsilon = 4$ to find the hidden linear relationship $R2$ with optimal w, b and ζ. Record the computational time and compute the corresponding MSE.

(iii) (5 points) Visualize $R1$ and $R2$ with all data points and the tubes with radius $\epsilon = 4$ for both. Also, mark those support vectors for $R2$.

(iv) (5 points) Table the MSE value, the number of data points outside the tube, and the computational time of $R1$ and $R2$ for comparisons. Discuss the advantages and disadvantages of applying the commonly used (least-squares) linear regression model and support vector regression model for analyzing this dataset.

(v) (20 points) Repeat tasks (i), (ii), (iii) and (iv) to uncover the hidden linear relations $R3$ and $R4$ embedded in the dataset "data2-ex3-s25.csv".

(vi) (5 points) Combining your results of (iv) and (v), list at least five lessons you learned from doing this exercise.

(6) For dataset "data3-ex3-s25.csv", perform the following tasks:

(i) (10 points) Use the LSSVR model (page 64 of Lecture 5) to find the hidden linear relationship $R5$ with optimal w, b and ζ. Determine and report the optimal regularization parameter C and ϵ. Please be sure to justify your parameter selections with reasons (in terms of MSE, data points outside the

tube and etc.). Record the number of support vectors, the computational time and the corresponding MSE.

(ii) (10 points) Use the DKLSSVR model (page 78 of Lecture 5) with the Polynomial kernel (page 43 of Lecture 5) and $\epsilon = 0.3$ to find the "best possible" hidden relationship $R6$. Record the number of support vectors, the computational time and the corresponding MSE. You certainly have to test and analyze many parameters to reach "best possible" outcomes.

(iii) (10 points) Use the DKLSSVR model (page 78 of Lecture 5) with the Gaussian kernel (page 45 of Lecture 5) and $\epsilon = 0.3$ to find the "best possible" hidden relationship $R7$. Record the number of support vectors, the computational time and the corresponding MSE. You certainly have to test and analyze many parameters to reach "best possible" outcomes.

(iii) (10 points) Visualize and discuss the results of $R5$, $R6$ and $R7$ with all data points and tubes with the radius $\epsilon = 0.3$ for each case.

(iv) (5 points) list at least five lessons you learned from doing this exercise.

(7) (10 points) Derive the Dual Soft Support Vector Regression (DLSSVR) model (page 70 of Lecture 5, the lower one) from the Linear Soft Support Vector Regression (LSSVR) Model (page 66 of Lecture 5). Do it step by step to make sure that you know how to get explicitly the dual variables, objective function, and linear constraints.

Report Format

(1) Title and Team Members

(2) Background and Objectives

(3) Work and Results

(4) Graphic Display of the Best Results

(5) Analysis and Discussions

(6) Attachment of Computer Program/Codes, Inputs and Outputs