MATH 304 - Numerical Analysis and Optimization Project

Issued: Jan.25th, 2021

Due : Feb.10th, 2021 (21:00 Beijing Time)

Please submit the PDF file and M files of your solution before the due time.

Topic: Least Squares Regression

Tasks:

Build a polynomial model to fit the given data sets.

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2 x^2 + a_1 x + a_0$$

When solving the model coefficients, you can consider to add a penalty term into the cost function by using L2-norm regularization. In this case, the cost function should be in the form of $||A \cdot \alpha - B||_2^2 + \lambda \cdot ||\alpha||_2^2$. Write your MATLAB code to build the over-determined linear equations that take the regularization term into account. Solve the model coefficients by using the backslash "\" in MATALB.

(1) Try different models (n=1, 2, 3, ... , 9) without regularization (λ=0) on the training data set of "SmallData.mat", and test the error on the "TestData.mat".

  (a) Fill in the following table to show the training error and test error for each model. The error means average squares error (i.e., normalized by the sample number).

Model	N=1	N=2	N=3	N=4	N=5	N=6	N=7	N=8	N=9
Training Error									
Test Error									

  (b) Plot all data (training data and test data with different colors);

  (c) Plot two fitted models in the same figure of (b) (the model with smallest training error and the model with smallest test error. if they are the same model, only plot that model);

  (d) Print the model coefficients in the task (c).

(2) Repeat the task (1) by using the training data set of "LargeData.mat", and show the results.

(3) Use the model (n=9) in the task (1) with different regularization weights ($\lambda = 10^{-6}$, $\lambda = 10^{-3}$, $\lambda = 1$, $\lambda = 10^3$, and $\lambda = 10^6$).
   (a) Fill in the following table, and repeat the tasks (b) (c) (d) in the task (1).

Weight	$\lambda = 10^{-6}$	$\lambda = 10^{-3}$	$\lambda = 1$	$\lambda = 10^3$	$\lambda = 10^6$
Training Error					
Test Error					

(4) Use the strategy of 5-fold cross validation to choose the best regularization weights from the range ($\lambda = 10^{-6}$, $\lambda = 10^{-3}$, $\lambda = 1$, $\lambda = 10^3$, and $\lambda = 10^6$) for the model (n=9) with the training data set of "LargeData.mat".
   (a) Show the average validation error for each regularization weights. (4 folds are for training, and the remained 1-fold is for validation.)
   (b) Using the best regularization weight with smallest validation error in (a), train the model again with all training data.
   (c) Print the model coefficients in (b), the best regularization weight, and the test error.

Submission:

(1) The Matlab code files (.m files) to be compressed as "MATH304-YourName-Proj-Code.rar or zip":
   (a) The function code of least squares regression is saved to a file named by "LSR-YourNetId.m"; (**3 marks**)
   (b) The main code for task 1, task 2 (**3 marks**), task 3 (**2 marks**), and task 4 (**2 marks**) are saved by the file names of "Main1-YourNetId.m", "Main2-YourNetId.m", "Main3-YourNetId", and "Main4-YourNetId", respectively.
   (c) If you have any other functions, please save them separately by "FunctionName-YourNetId.m"

(2) One final report "MATH304-YourName-Proj-Report.pdf" with the template. (**15 marks**)