CS-584 – Assignment 1 (5%)

Parametric regression

Due by: February 18, 2016

Assignment Specifications

In this assignment you need to implement techniques for parametric regression. Sample data files for this assignment are available on the cs584 website. Use both synthetic and real datasets. It is essential that you evaluate the performance of each algorithm you implement. To test the performance of your implementation use 10 fold cross validation. The grade for this assignment will be based in part on the performance of your implementation and on the thoroughness of your evaluation. Unless otherwise specified you need to implement the algorithms yourself and not use the ready made Python function to do this. You may use all other Python functions.

1. Single variable regression:
   (a) Load each of the single feature data sets and plot the data to get an idea of the complexity of the problem.
   (b) Fit a linear model to the data. Compute the training and testing error, and compare them. Plot the regression model you obtain on top of the test data.
   (c) Compare the results you obtain to the one obtained by the ready made Python function.
   (d) Test different polynomial models on different subsets of the data. Choose a single polynomial model and justify your selection. Repeat the evaluation procedure as above.
   (e) Reduce the amount of training data you use and observe the effect of performance of the linear and polynomial models.
2. Multivariate regression:
   (a) Load the multiple feature data sets, and map them to a higher dimensional feature space using combinations of features.
   (b) Perform linear regression in the higher dimensional space. Evaluate different mappings in terms of the testing error they produce. Choose a single model and justify your selection. Compute the mean square error when using the training set and when using the testing set.
   (c) Solve the regression problem using an explicit solution and using an iterative solution. Compare the two solutions.
   (d) Using a Gaussian kernel function, solve the dual linear regression problem. Compare the time performance and accuracy of the dual regression to the primal regression problem you solved earlier.

General comments

- Implement the assignment using Python.
- Write your code in a modular way using functions and make sure to document it.
- Do not include in the submission large datasets that were provided by us.
- In plotting unsorted points, do not connect them with line segments.
- Do not include repetitive results. Show only results that have a purpose.

Electronic Submission Instructions

Please follow the following submission procedure:

1. Direct all questions/comments regarding the assignment to: *cs*584@*cs.iit.edu*
2. On or before the due date upload your submission to blackboard. Attach to it a ZIP file containing the following:

   - Report: prepared as a PDF file. The report should contain a summary of program design issues, description of specific problems you faced and the way in which you solved them, and sample input/output results (text/graphic). The report needs to be sufficiently detailed.
   - Code: all the source code files that are necessary to execute your program. Please do not submit data files of saved results.

   Note: we must be able to view your report and execute your program in order to grade it
3. The organization of the submitted material should be as follows:

   - Create a directory called: `first_last_ass#` where "first"/"last" is your name and "#" is the assignment number.
   - Inside this directory create four sub-directories called: `code`, `report`, `data`. Place in these directories the files you need to submit.
4. Do not submit a paper copy of your report. You will be contacted by email if some material is missing or if you will need to meet with the TA.
5. If you are late in the submission, submit it as soon as you have it. "late days" will be determined by your submission date.