Computer Based Test 1 - Machine Learning (extended)

Linear Regression by Least Squares

Notes: This is a Computer Based Test. You will need a computer and MATLAB to solve it. This test is compulsory for students registered for the Machine Learning (extended) module. It contributes 1/3rd towards the 20% allocated for Computer Based continuous assessment marks. Recall that ML Extended has 40% continuous assessment, of which half (i.e. 20%) is Computer Based Test.

Students registered for the Machine Learning module are encouraged to try it, however they do not need to submit solution and it will not be counted towards their grade.

Submission Deadline: 23.59 on Friday 21st October, 2016 (end of week 4). Late submission will carry a penalty of 10% loss per day. Submissions more than 72 hours delay will not be accepted.

Deliverables: Submit a single PDF file on Canvas containing solutions for all the tasks. The PDF file should contain the MATLAB code developed, with due credits (to any borrowed part from the book) and summary of your findings including any graphical plots for each task. The first page should include the student name and ID number.

How to submit?: via Canvas by uploading a file.

Plagiarism Policy: Any work submitted in your name should be YOUR work, except where due credit is cited.

https://intranet.birmingham.ac.uk/as/studentservices/conduct/plagiarism/index.aspx

https://intranet.birmingham.ac.uk/as/studentservices/conduct/plagiarism/guidance-students.aspx

http://www.birmingham.ac.uk/Documents/university/legal/plagiarism.pdf

Marking Scheme: The submitted work will be graded on the basis of completeness, correctness, and neatness.

Any questions?

Please ask in good time, not in the last minutes before deadline! The best time to ask questions is during the weekly office hours on Tuesdays 9.30am-11am. Alternatively, you may ask or get appointment via email (k.m.rajpoot@cs.bham.ac.uk).

Test Tasks

Download the MATLAB code and data associated with this test: **CBT1.zip** (https://canvas.bham.ac.uk/files/3264470/download?download_frd=1). This includes Olympics data (**olympics.mat**), linear least square fitting and prediction code (**olympicpred.m**), polynomial least squares fitting code (**olymppoly.m**), and cross-validation experiment code on random data (**cv_demo.m**). Run these MATLAB code files and try to fully understand their operation by associating MATLAB commands with results and plots obtained. This will provide you useful insights and MATLAB background to solve the tasks below.

Solve the following two tasks:

T1: Load the Olympic men's 400m data (stored in the file **olympics.mat**). Fit 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} order polynomial functions to this data. Use 10-fold cross-validation to choose the "best" polynomial order between 1 and 4 which has lowest error. What is the "best" model based on average cross-validation loss?

T2: Using 5-fold cross-validation, find the value of λ that gives the "best" predictive performance on the Olympic men's 100m data (stored in the file **olympics.mat**) for (a) 1^{st} order polynomial (i.e. the standard linear model) and (b) 4^{th} order polynomial.