

Midterm

Machine Learning COMS 4771, Spring 2014, Itsik Pe'er

Assigned: March 22nd

Due: Friday, March 26th 1:10pm

Submission: Your submission folder on Courseworks. Submit a zipped folder for

Midterm.Problem01 or Midterm.Problem02

You are provided with a dataset of $N=10000$ datapoints in the MatLab file `Data.mat` that can be loaded into MatLab by the `load('Data.mat');` command. The file includes three variables:

X : $N \times 9$ matrix of real values, interpreted as vectors that are *input* datapoints in $D=9$ dimensions

y_{real} : $N \times 1$ matrix of real values, interpreted as *real output* datapoints in one dimension

y_{int} : $N \times 1$ matrix of $\{1,2\}$ values, interpreted as *integer output* datapoints in one dimension

Implement a MatLab function `predict` that receives two inputs, and returns one output, all formatted as detailed above. `predict` solves one of two prediction problems (you will not be graded for solving both) below:

Problem 01: Predict real output given input and integer output

Problem 02: Predict integer output given input and real output

You will be evaluated (50% of grade) on testing data drawn from the same distribution as the training data (regression error or classification error, respectively). If `predict.m` uses any other MatLab code or libraries, they should be included in your submission folder.

Besides `predict.m` you are further required to submit an explanation of your solution (50% of the grade). This should be submitted either electronically in `explain.pdf` in the submission folder or as a hard copy (written or printed) before the deadline. You are free to use any method discussed in the course thus far. Cite any references you use. If your explanation relies on any computation, e.g. failed attempts, or statistics extracted from the data, include those in the submission folder and refer to them in your explanation. These computations must be in MatLab and must be implementations of material covered in class.

Recall, that this is an exam. You are not allowed to discuss the exam with any person besides the course staff. You are not allowed to post any message about it on any forum, user community or social media outlet, with the exception of private messages to the course staff on Piazza. Course staff will be on call 9am-6pm to respond, per the course calendar.

The exam is open book in the sense that you may use any paper material that was generated by others and you obtained before the start of the midterm. You may read any electronic material and run any MatLab code that implements material covered in class, as long as these had been saved in computer folders that you own before the start of the midterm. You may search for additional electronic information at the course courseworks site or in `mathworks.com`, but not beyond that.

Good luck!