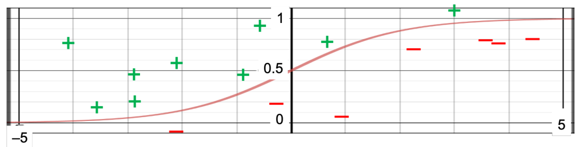
CS3244 Exam 1: Part 3

28 Sep 2020

Please do not turn to the next page until you are told to do so by your proctor.

- This exam part is worth **25** marks out of a **75** mark total for all three parts.
- This exam part is estimated to take you about **25** minutes to complete.
- This exam part has a total of **8** questions.
- This exam part has only *text response* (essay) questions and *worked response* (show your intermediate workings / calculations) questions. Be mindful of the estimated time for a response; longer answers may incur penalties and cost you extra time that you could use to answer other questions.
- You can visit `http://www.comp.nus.edu.sg/~cs3244/2010/e1.part3.html` to reach the entry form for this survey if you lose your browser window.
- Do remember that you will need to key in and submit your answers to the according assessment system as designated by your proctor or by exam central.

[Questions 1–2] The below image purportely shows how a Logistic Regression classifer boundary looks like (A version of this image did actually appear on a fairly famous data science website).



1. (Text Response; 3 marks) Describe why this is an incorrect representation of how Logistic Regression functions.

2. (Text Response; 1 mark) Describe one aspect of this image which is representative of Logistic Regression.

[Questions 3–5] You have been given n datasets — $D_1, D_2, ..., D_n$ — and a supervised learner L to learn h. L is to learn a single variable function $h : x \rightarrow y$, where $x, y \in \mathbb{R}$. We want to use the learner and different datasets to learn and predict using the average hypothesis: $\bar{h}(x)$.

The relevant classes are defined below using pseudocode:

```
1  class Hypothesis {
2      /* This class represents a hypothesis - h(x) - returned when a learner learns from a dataset.
3         We only show the relevant function singatures and variables.
4      */
5      int predict(int x) {
6          // Computes the prediction for input x (integer) and returns h(x) an integer
7      }
8  }
9
10 class Learner {
11     /* This class represents a learner based on some learning algorithm (eg. SGD) and hypothesis class.
12     */
13     Hypothesis learn(Dataset D) {
14         /* Takes a dataset D and returns an object of class Hypothesis.
15            You can assume the hypothesis returned is the one with least training error.
16         */
17     }
18 }
19
20 class AverageHypothesis extends Hypothesis {
21     /* This class represents the average hypothesis and will be the one you are completing.
22     */
23
24     hypotheses = [] //List/Array to be populated
25
26     void initialize(Dataset D1, Dataset D2, ..., Dataset Dn) {
27         //This function will populate the hypotheses list
28     }
29
30     int predict(int x) {
31         //This function returns the prediction of the average hypotheses on input x
32     }
33 }
```

3. (Worked Response; 3 marks) Code the INITIALIZE function (Line 26) in the AVERAGEHYPOTHESIS class. Assume that the other classes' functions are complete. Clear pseudocode is sufficient.

4. (Worked Response; 3 marks) Code the PREDICT function (Line 30) in the AVERAGEHYPOTHESIS class. Assume that the other classes' functions are complete. Clear pseudocode is sufficient.

5. Now assume that the learner L learns a function of the hypothesis class $y = mx + b$, and you are given 3 2-point datasets in the form of the (x, y) values below:

$$D1 : (1,3),(2,4)$$
$$D2 : (1,5),(2,1)$$
$$D3 : (1,2),(2,9)$$

   (Worked Response; 4 marks) Describe the output of the call to your INITIALIZE function, and subsequently, what your PREDICT function returns for $x = 3$. Assume the LEARN function runs until the best hypothesis is obtained given the complexity of the hypothesis class.

   Show relevant intermediate return values for full credit.

6. (Text Response; 4 marks) The expressive power of linear models is low compared against other learners that capture both linear and non-linear relationships natively.

   Why then are linear models still relevant in the real-world? Justify your response.

7. (Text Response; 3 marks) We learned about **Laplace Smoothing** and showed how it solves zero values in Naïve Bayes. Answer whether you would or would not apply Laplacian smoothing for skewed data. Justify your response.

8. (Text Response; 4 marks) Say that you know stochastic noise is very high in a dataset you have been assigned to conduct supervised learning on. State one condition for why you would use a high variance model, and one condition why you would not.

   Justify your answer for full credit.

This marks the end of this part of the exam.
These is no additional material beyond this point.