QUIZ 2

purdue university · cs 51500

David Gleich

matrix computations

June 27, 2023

Purdue Honor Code

The purpose of the Purdue University academic community is to search for truth and to endeavor to communicate with each other. Self-discipline and a sense of social obligation within each individual are necessary for the fulfillment of these goals. It is the responsibility of all Purdue students to live by this code, not out of fear of the consequences of its violation, but out of personal self-respect. As human beings we are obliged to conduct ourselves with high integrity. As members of the civil community we have to conduct ourselves as responsible citizens in accordance with the rules and regulations governing all residents of the state of Indiana and of the local community. As members of the Purdue University community, we have the responsibility to observe all University regulations.

To foster a climate of trust and high standards of academic achievement, Purdue University is committed to cultivating academic integrity and expects students to exhibit the highest standards of honor in their scholastic endeavors. Academic integrity is essential to the success of Purdue University's mission. As members of the academic community, our foremost interest is toward achieving noble educational goals and our foremost responsibility is to ensure that academic honesty prevails.

Exam Rules

AP_____This quiz is a *open to all class and internet resources* except talking with people.

AP_____The contents of this quiz are protected by copyright. Posting any piece of this quiz – including excepts that may be subject to fair-use rights – to any database, online resource, electronic medium, or other similar repository will be considered academic dishonesty (see below). Exams may be individually watermarked to identify violations.

AP_____You may not consult with anyone other than the professor and the TA.

AP_____Any behavior consistent with academic dishonesty (i.e. cheating) *will not be tolerated and may result in a 0 even a failing grade in the class*.

AP You have until the Gradescope or Blackboard submission closes to com plete and submit the exam. No exceptions.

Write your name, PUID, and sign below to indicate you agree with the statement:

The remainder of this exam represents my own work.

(Your Name) Aditya Patel (PUID) 0035573514

Aditya Patel
(Signature)

Problem 1 *(10 points)*

To explain how accurate floating point values were, we started using the reference point from earth to the sun. Using as a reference point 1 kilometer, identify between 5 and 10 other items that would help contextualize using $2^{52} \approx 10^{16}$ values to denote ”one kilometer”.

Letting 1 km scale to 10^{16}:

- 10^{15}: (0.1 km) An American football field is about 100 m, or 0.1 km
- 10^{14}: (0.01 km) A Tennis court is about 24 m, or 0.024 km
- 10^{13}: (0.001 km) An average human is about 1.7 m, or 0.0017 km
- 10^{12}: (0.0001 km) A wooden pencil is about 19 cm long, or 0.00019 KM
- 10^{10}: (0.000001 km) A black ant is about 4 mm long, or 0.000004 km

Problem 2 *(10 points)*

Answer true or false, no justification needed, but answers that differ from what we consider the "best answer" may still yield points if you have a compelling justification.

(2 points) Assuming computational cost and memory are not an issue, floating point values should always been used in every scenario.

False.

(2 points) We don't get floating point errors when sorting a list of integers.

True. Integer comparisons are done as fixed-point numbers.

(2 points) Counting the number of events that occur and using that to estimate a probability is a likely instance of a floating point error.

False. Number of events is an integer, and the occurring number of events is also an integer. An integer divided by an integer would not result in floating point error.

(2 points) If x is a value that is around 10000 and y is a value that is around 0.01. Then computing $y/(x - y)$ will have a large floating point error.

True. Since the problem does not state the scale at which the numbers are 'around', there are an infinitesimal number of values that can be counted as 'around' 10000 and 0.01.

(2 points) The value of unit roundoff for the Float32 system is 2^{-52}.

False. The 2^{-52} is the unit roundoff for Float64

Problem 3 *(20 points)*

We saw a video that showed that when adding 3 numbers, the order from smallest to largest gave the lowest error. What happens if we multiply numbers instead of adding them? Repeat this style of analysis for multiplication.

$$Consider: x > y > z > 0, and\ |\mu_1| = |\mu_2| \leq u$$

$$\boldsymbol{For}: (\boldsymbol{x} * \boldsymbol{y}) * \boldsymbol{z}:$$

$x * y = (x * y)(1 + \mu_1),$
$(x * y) * z = \big((x * y)(1 + \mu_1) * z\big)(1 + \mu_2)$
$\Rightarrow \big(x * y * z + \mu_1(x * y * z)\big)(1 + \mu_2)$
$\Rightarrow (x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2 + (x * y * z)\mu_1\mu_2$
$\Rightarrow (x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2 + 0$
$\Rightarrow \boldsymbol{(x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2}$

$$\boldsymbol{For}: \boldsymbol{x} * (\boldsymbol{y} * \boldsymbol{z})$$

$y * z = (y * z)(1 + \mu_1)$
$x * (y * z) = \big((y * z)(1 + \mu_1) * x\big)(1 + \mu_2)$
$\Rightarrow \big((y * z + (y * z)\mu_1) * x\big)(1 + \mu_2)$
$\Rightarrow (x * (y * z)) + (x * (y * z))\mu_1)(1 + \mu_2)$
$\Rightarrow (x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2 + (x * y * z)\mu_1\mu_2$
$\Rightarrow (x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2 + 0$
$\Rightarrow \boldsymbol{(x * y * z) + (x * y * z)\mu_1 + (x * y * z)\mu_2}$

We can see above that the error terms are the same for both multiplication methods, which means there is no difference in error in the way we multiply.

Problem 4 *(10 points)*

(8 points) Consider the numerical computation involved in training a speech recognition system that uses a deep neural network. This involves files of audio and labels that have accurate transcriptions. Articulate as many elements of the numerical computation involved in terms of the things we have learned in class.

Data: The data is the input given to the numerical computation. In this case, the data is the labeled set of audio transcriptions and the audio files.

Model: The model is the mathematical principle behind the computation. In this case, the model would be the DNN.

Algorithm: The algorithm is the actual code and programming done to implement the model. In this case, the algorithm isn't explicitly stated, but based on previous knowledge, DNNs can be implemented in Python using packages such as PyTorch and TensorFlow.

(2 points) Suppose that the model is found to perform poorly for foreign accents. Do you think the data, model, or algorithm is likely to be the problem?

Based on the information provided:

- If the data has been correctly labeled and the transcription are accurate, which is given they are, there is no input issue.
- If the model was incorrectly selected, then the model would perform poorly for all inputs, regardless of accents.

Therefore, there must be an issue in the implementation of the model, which is the algorithm.