Graded Exercise 1

WRITE YOUR NAME HERE
CSC131 The Beauty & Joy of Computing

14 September 2018

1. Suppose that we have records of the performance of many athletes in 10km (about 6.2 miles) and marathon (26.2 mile) races.

   We can plot this data on a graph by interpreting the best time of an athlete at the 10km distance as a horizontal (x) coordinate and the best time at the marathon distance as vertical (y) coordinate.

   For each athlete, we have a point (x, y) on our graph, where x is the athlete's 10km time and y is the athlete's marathon time. In particular, for the i^{th} athlete we have a point (x_i, y_i)

   We can use this data to build a model. Our model will enable us to predict a runner's time in the marathon, given the runner's time in the 10km race.

   We will suppose that there are numbers m and b such that $y = m \cdot x + b$ gives a good (but not perfect) estimate of the runner's performance. This equation describes a line. We will seek values of m and b that give us the best line.

   If we know x_i (the 10km time of the i^{th} athlete), we can compute $y = m \cdot x_i + b$ to get a prediction of the athlete's marathon time. In this case, it is very unlikely that all of our points lie on a straight line, so there will in general be a difference between the actual marathon time y_i and the predicted marathon time y.

   The difference $(y_i - (m \cdot x_i + b))$ is an error—it is the difference between the measured and predicted performance of a athlete. Let's call the error for the i^{th} athlete e_i.

   To get the best line (and therefore the best predictor), we will minimize the sums of the squares of all of the errors.

   Why will we sum the squares of the errors rather than sum just the errors?

   That is, why compute $e_0^2 + e_1^2 + e_2^2 + e_3^2 + \ldots$ rather than $e_0 + e_1 + e_2 + e_3 + \ldots$?

WRITE YOUR ANSWER HERE.

2. There is an algorithm for solving the traveling salesman problem. It is a simple algorithm. A programmer can translate this algorithm into a short and correct program.

   The solution of the traveling salesman problem is easy if the traveler's circuit includes only a few stops. It is beyond our reach if the journey includes a hundred stops.

   If we have simple algorithm and correct program, why can we not get a solution to the problem?

WRITE YOUR ANSWER HERE.

3. Mathematicians can solve some problems directly. If you give me the current temperature measured in degrees Farenheit, I can give you the temperature in degrees centigrade: $C = (5/9) \cdot (F - 32)$.

   Mathematicians solve other problems iteratively. They make some initial guess or estimate. Then they make a better guess or estimate. They produce better and better estimates until they have one that they call "good enough."

   Here's an example: Suppose that we have guessed that the square root of 2 is x. A better guess is the average of x and $2/x$.

   You can measure your error at each step: $error = \mid 2 - x^2 \mid$.

   You can decide what "good enough" means. In some circumstances, you might be willing to tolerate an error of 0.1. In other circumstances, you might need an estimate that is accurate to within 0.000001 of the true value.

   Begin by guessing that the square root of two is $x = 1$. Compute a better guess, and then still better guess, and then a guess that is even closer to the true value of $\sqrt{2}$. At each step, replace the value of x with $(x+2/x)/2$. Use a calculator or a spreadsheet program.

WRITE YOUR ANSWER HERE.

4. Gradient descent is a method of successive approximation. Describe this algorithm with an analogy that involves a hiker.

WRITE YOUR ANSWER HERE.

5. A team of data scientists have obtained the purchasing histories of many customers of a large retailer. The data scientists want to create a model of the customers' behavior that will enable them to predict future customer choices. Will they put all of the data that is available to them into the training set? Explain.

WRITE YOUR ANSWER HERE.

6. What is a stochastic method?

WRITE YOUR ANSWER HERE.

7. Peter took an interest in how scientists can use machine learning to increase their understanding of the data that they generate in their experiments. He named a scientific instrument that generates enormous amounts of data. Which instrument?

WRITE YOUR ANSWER HERE.

8. Easton wrote about a machine learning program called AARON. The program has a long history. How long? What is the purpose of the program?

WRITE YOUR ANSWER HERE.

9. Maddy expressed concern that technology will erode the quality of human relationships. How might that happen?

WRITE YOUR ANSWER HERE.

10. Jakob found a report on efforts to detect signs of Alzheimer's disease. How does the technology do this?

WRITE YOUR ANSWER HERE.

11. Rodrigo wrote about Content Clarifier. This technology uses the IBM Watson system. It helps people with what kinds of disabilities? How does it help them?

WRITE YOUR ANSWER HERE.

12. Induction and deduction are two methods of reasoning. They work in opposite directions. How so?

    Which method does Sherlock Holmes use to identify the person who committed a crime? Which method might a mathematician use to prove that the sum of the first n positive integers is $n(n+1)/2$? (For example, when $n = 3$ it is the case that $1 + 2 + 3 = 3(3 + 1)/2$.)

WRITE YOUR ANSWER HERE.

13. Nicole discovered efforts to enable physicians to better diagnose mental illnesses. Which illnesses? Using what kinds of inputs to the computer?

WRITE YOUR ANSWER HERE.

14. Marcellus read about precision medicine. What is precision medicine?

WRITE YOUR ANSWER HERE.

15. William wrote about a project at a German university. People there trained a robot to imitate the way children build with a popular product whose origins are in neighboring Denmark. Say more about the goals and methods of this project.

WRITE YOUR ANSWER HERE.

16. What is the DeepQA project at IBM? Tiff included some information about this use of the Watson system in her report.

WRITE YOUR ANSWER HERE.

17. Koichi commented on an ethical dilemma that the people who are developing an important application of machine learning will have to resolve. What is that ethical dilemma?

WRITE YOUR ANSWER HERE.

18. Marcellus described a wearable device that is linked to smart phone. What is the medical use of this technology?

WRITE YOUR ANSWER HERE.

19. Many of the courses that you recommended listed similar mathematical prerequisites. What kinds of mathematics should someone who wants to understand machine learning algorithms study?

WRITE YOUR ANSWER HERE.

20. What are two programming languages that you saw listed for use in one or more of the courses that you or your classmates recommended?

WRITE YOUR ANSWER HERE.