Graded Exercise 1

Answer Key
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 $e_0 + e_1 + e_2 + e_3 + \ldots$?

In a sum of errors, positive and negative errors could cancel one another. The sum could be zero even if the magnitudes of many errors are large.

The square of a number is a non-negative number. Summing the squares of the errors yields a sum of non-negative values. This is better measure of the predictive power of the model.

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?

The algorithm requires a number of arithmetic operations that rises quickly as the number of stops on the tour increases. If the number of stops is large, the number of arithmetic operations needed to select the optimal route is so large that even the fastest computer cannot produce the result in a human lifetime.

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.

The value of $\sqrt{2}$, to an accuracy of four digits beyond the decimal point, is 1.4142 It takes just a few refinements of an initial guess of 1.0 to get to that value.

Iteration #	**Estimate**
0	1.0000
1	1.5000
2	1.4166
3	1.4142

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

A hiker in hilly terrain and in a thick fog wants to find a way to the bottom of the deepest valley. Because of the fog, the hiker cannot see far, but the hiker can, by extending a foot out a few centimeters to the front, to the back, and to each side, feel the shape of the nearby ground. In this way, the hiker can determine the direction in which the ground slopes most steeply downward. Moving in that direction will take the hiker down the mountain.

There is a risk that the hiker will wander into a depression that is high above the bottom of the deepest valley. While standing in this depression, the hiker senses that the ground slopes upward in every direction.

If you choose to continue your study of algorithms, you will find methods of overcoming this problem.

5. A team of data scientists has 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.

The scientists will not put all available data in the training set. They will reserve a fraction of the available data to use in testing the model that they produce.

If the scientists trained the model with all available data, they would have no means to evaluate the quality of their model.

6. What is a stochastic method?

"Stochastic" means random.

Some algorithms make random guesses. For example, suppose that you have a shape S within a rectangular region R. Suppose furthermore that you have a method of determining whether or not a given point lies inside the shape S and a method of selecting points in the containing region R at random. You know the area of R and want to know the area of S. By testing a large number of random points and counting how many are inside S, you could estimate the area of S.

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?

The Large Hadron Collider is an instrument that measures the trajectories of colliding sub-atomic particles.

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?

Harold Cohen has been working on AARON for more than forty years. AARON produces art.

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

If people can converse with robots, and especially if those robots can provide a sense of emotional intimacy, they might withdraw from human society.

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

The software analyzes the voices of patients.

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?

The software helps people with dementia and autism. It translates figures of speech into plainer and simpler language.

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 $1 + 2 + 3 = 3(3 + 1)/2$.)

Sherlock Holmes deduces the identity of the criminal. The detective applies general knowledge (perhaps the speed with which mud dries) to find specific facts (the person responsible for tracking mud into the kitchen). Detectives need good powers of deduction.

Mathematicians can begin with specific facts and find a general rule by the method of induction. In our class, I proved that the number of people who have shaken hands with an odd number of people is even. I moved from small, specific cases (the statement was true before the invention of the handshake and it was true after the first two people—maybe Adam and Eve—shook hands) to the ever bigger cases (all of the handshakes exchanged by all of the people now alive and all of the people who ever lived).

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

One project is applying machine learning to the interpretation of many symptoms of post-traumatic stress. Another project combines machine learning and medical imaging technologies in an effort to understand the brains of people afflicted with schizophrenia.

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

Precision medicine means the selection of treatments that are tailored to the individual patient. It offers the prospect of using a patient's unique genetic makeup to identify an effective treatment. Without this technology, physicians might try several treatments before finding the right one, and so expose patients to greater risk and costs.

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.

A group at Freiburg University of Mining and Technology trained a robot to build with Legos. They did they as part of an effort to build machines that can learn by watching people and imitating their actions.

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

DeepQA can read a doctor's notes and other non-standardized texts.

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?

The computer program that controls a self-driving car might, in a case where it cannot prevent an accident, have to choose between an action that injures one person and an action that injures another person.

The development of self-driving cars will require answers to questions about how to assign responsiblity when accidents occur.

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

The device has the potential to detect disease in the heart.

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?

The descriptions of several courses recommended a foundation in calculus and linear algebra.

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?

Several courses included exercises with the R and Python programming languages. You might have seen other programming languages listed in the descriptions of courses.