University of Illinois at Urbana-Champaign
ECE101 Exam II
Wednesday 30 March 2022

Name:

SOLUTIONS IN RED
SOLVE YOURSELF FIRST!

Net ID:

- Be sure that your exam booklet has SEVEN pages.

- This is a closed book exam. You may NOT use a calculator.

- You are allowed one 8.5×11” sheet for any notes (both sides).

- Absolutely no interaction between students is allowed.

- Show all work, and clearly indicate any assumptions that you make.

- Don’t panic, and good luck!

<table>
<thead>
<tr>
<th>Problem</th>
<th>Points</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Problem 2</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Problem 3</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Problem 4</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Problem 5</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>
Problem 1 (21 points): Information and Encryption

A. **(7 points)** A prominent company has developed an online recommendation service for restaurants in major US metropolitan areas. Based on users’ previous reviews and frequency of visits, the company suggests restaurants that each customer might enjoy visiting. Customers make use of the recommendations both in their home cities and while traveling in one of the other areas covered by the service.

The company is now thinking of expanding into a new market. Specifically, the company wants to use its data to provide suggestions for restauranteurs in areas that have yet to develop the full diversity of cuisines available in many large cities. In two or three sentences, explain how the company can make use of its data to make suggestions for new restaurants in City C. Note that the company does not have customers (yet) who live in City C, but customers from other cities have visited City C and rated restaurants in the past using the company’s app.

Using “eigencustomers” together with existing rating data from City C, identify types of restaurants that customers normally visit but did not visit while in City C.

Presumably, customers did not visit that type of restaurant because they were absent from City C, so suggest developing the missing types.

B. **(7 points)** Now imagine that the company acquires another company that has been collecting similar data for several years on City C’s citizens. How can the company combine these new data with its own data to improve its recommendations for new restaurants? Explain in one or two sentences.

If City C’s citizens frequented a type of restaurant when traveling but did not go to that type while in City C, developing new restaurants of that type in City C seems promising.

C. **(7 points)** Cryptographic keys are usually too large for humans to remember (256 or 512 Bytes), so they are stored by using a short password (or, better, a pass phrase) to encrypt the key. When the key is needed, the owner types the password, which allows the key to be decrypted and then used.

An employee of a cloud provider suggests using each user’s account password to protect a key stored with the company. The key can then be used to encrypt all of the data stored for that user in the company’s cloud. In one or two sentences, explain a major drawback of the employee’s idea.

The encryption adds little or nothing over the password itself in this case, since an attacker who obtains a user’s password also has immediate access to their key.
Problem 2 (21 points): Machine Learning Pros and Cons

Walcorp is thinking about replacing most of its human employees with machine learning.

Specifically, based on the purchasing habits of its many customers over the last two decades, Walcorp plans to develop a weekly delivery of groceries and other household items personalized to each customer. These deliveries can then be grouped by the truckload and routes planned to optimize delivery efficiency by a new fleet of autonomous trucks. In the week preceding a given delivery, a household can visit Walcorp online to tailor that week’s delivery, modifying or removing regular items and adding any exceptional items. Most stores will be closed, allowing Walcorp to downsize most of its employees.

Given what you know about machine learning, describe what you think are the two most important advantages that Walcorp might see in its plan. Also describe what you think is the biggest disadvantage from the perspective of Walcorp. Explain each in a sentence or two.

ADVANTAGE # 1 (7 points): cost savings from employee training, salary, and benefits

_____________________________________________________________________________________
_____________________________________________________________________________________
_____________________________________________________________________________________

ADVANTAGE # 2 (7 points): reduced need for floor space, both due to the absence of

_____________________________________________________________________________________

customers and to the more predictable inventory requirements to meet weekly demand
_____________________________________________________________________________________
_____________________________________________________________________________________

DISADVANTAGE (7 points): no “drop-in” customer sales; also, any customer who needs

_____________________________________________________________________________________

an unusual item sooner than their weekly delivery will shop elsewhere
_____________________________________________________________________________________
_____________________________________________________________________________________
Problem 3 (20 points): Numerics and Quantification

The chart below is based on data provided publicly by the US CDC. The dotted line shows a linear fit to the average weight of males and females as a function of age.

A. (5 points) Do you think that using separate linear fits (separate lines) for males and females would give better results, or worse results? CIRCLE EXACTLY ONE ANSWER.

1) Two lines give worse predictions than the joint line (shown).

2) Two lines give better predictions than the joint line (shown).

3) Using two lines is the same as using one line.

B. (5 points) Do you think that a single line gives a good fit for females (say within 5 kg for any age)? CIRCLE THE SINGLE BEST ANSWER.

1) Yes, a linear fit always produces accurate results.

2) Yes, a linear fit to a single data set follows the curve of the data precisely.

3) No, the data shown are not linear enough for any line to be so accurate.

4) No, linear fits are never accurate.
Problem 3, continued:

C. **(5 points)** Six web sites are mapped by your web crawler, producing the graph shown below. If the page rank algorithm is run on this graph, which node do you expect to have the highest rank (the highest reputation, most important, and so forth)? Explain your answer in one or two sentences. *Hint: Note that some nodes may seem to be tied, so explain how such a tie will be broken if you can.*

Node B should have the highest rank. While node C has the same number of incoming arcs, node B has an edge from node C (which is also large in-degree).

D. **(5 points)** In a sentence or two, explain why the function $R(x) = (997 \times x + 400) \mod 256$ is not a good choice for generating pseudo-random numbers. *Hint: you do not need to compute successive values of a sequence, just think about some of the problems that we discussed in class.*

The smallest bit in a sequence from $R(x)$ will always be the same, either odd or even.
Problem 4 (18 points): The Good, the Bad, and the Ugly

Realizing that most UIUC students carry their laptops around all day, that each laptop has a unique WiFi address, and that those laptops regularly ping the nearest router, thereby giving away their location, Prof. Victor Perkins has collected data tracking the location every student on campus throughout the day. Using login records, he has also associated the motion with each individual’s net id, which then gives him access to a wealth of personal information about each student. Prof. Perkins plans to train deep learning models using these data, then to make use of the models for whatever purpose he dreams up.

Help Prof. Perkins by coming up with TWO possible uses of the data to make predictions. You may choose any predictions that could reasonably be provided by the data. However, your suggestions must obey certain rules (see below). Describe each type of prediction in two or three sentences. Choose any two of the three below—leave the third blank.

A. (9 points) One purpose must be anonymous (not include any details about specific students).

Predict which study locations are most/least populated at different times of the day and on different days of the week.

B. (9 points) One purpose must make predictions about specific individuals—any type of prediction is acceptable, but they must be tuned to a given student’s behavior.

Predict how much time each student spends engaged in various activities on campus each day, then send these predictions to their academic advisor to provide insight on their behavior.

C. (9 points) The last purpose must be (at least borderline) illegal, unethical, or support illegal activities.

Sell access to “Where’s My Friend,” a student location prediction engine based on past data. For extra money, VIP customers also gain access to live tracking information!
Problem 5 (20 points): Understanding the Terms

For each term below, write the number corresponding to the correct definition from the list on the right. The definition chosen should correspond to the term’s use in lecture and lab. Note that not every definition will be used, as there are more definitions than there are terms.

7. A. authentication 1. a collection of movies played during prime viewing time—features, as they’re sometimes called

13. B. feature space 2. the use of a single key to encrypt and decrypt data sent over a communication channel

15. C. linear fit 3. an algorithm that ranks a set of web pages based on the number and organization of links (URLs in one page that point to other pages) between them

11. D. ML model training 4. the process of moving one’s data from web server to web server so that one’s activities are harder to monitor

8. E. net neutrality 5. organizing one’s papers into a single filing cabinet so that finding any paper is a linear process, simply starting at one end of the cabinet and looking at each paper in order

3. F. page rank 6. gathering information about a person’s activities

14. G. recommendation engine 7. association of a computer program with a person or a company; usually relies on cryptography

2. H. symmetric-key cryptography 8. the idea that Internet Service Providers and routers should not make decisions about Internet packets based on the purpose or people associated with the packets

6. I. surveillance 9. the process of verifying that an article presented as an antique is, in fact, not a fake

12. J. web crawling 10. an elite educational program in which the models used to advertise machine learning systems are trained to act

11. the use of example inputs and labels to develop a model that is able to produce the correct label for most or all of the example inputs

12. the process of exploring the World Wide Web by finding URLs embedded in a web page, following those URLs to find more web pages, and so forth

13. a mapping in which a set of things are represented as a set of numerical values, each of which can be thought of as a feature

14. an AI strategy that makes recommendations to a person based on that person’s past behavior and/or similarity to other persons

15. given a set of two-dimensional data (say X and Y), finding the line that minimizes the distance between the data points and the line