Study Questions for ECE 101 Exam 2

Spring 2022
Lectures 11 to 17

Table of Contents:

Questions only

`Q1` [L16] Explain the process of learning by example, and describe a situation in which a model learned by example might not reflect reality.

`Q2` [L16] Heavy farm equipment is a rare sight on most highways in the North Pacific Coast of the US (Washington, Oregon, Northern California). Explain why autonomous vehicles sold in these regions must still be trained with simulations including such equipment.

`Q3` [L16] A company wants to provide a service to locate humans who might potentially be in danger from forest fires—whether because they have not followed evacuation orders, were hiking, or some other reason. Give four reasons that using ML for such a task is superior to hiring humans.

`Q4` [L16] During a heavy rain, an autonomous vehicle approaches an intersection in which a vandal has cut down the stop sign with a hacksaw. Open-source maps mark the intersection as a 4-way stop. Unfortunately, the vehicle fails to stop and hits another vehicle. Who do you think should be held responsible for the accident? Do you think a human driver would have avoided the accident? Explain your answers.

`Q5` [L16] The New York City Public Library allows citizens of the city to borrow physical library materials. Recently, the library acquired a set of e-books, which it lends electronically. Your friend, an Urbana resident from Hawaii, has managed to obtain a library card by using a VPN to trick the library into believing that they live in Manhattan. Assuming that your friend's deceit is detected, should they be punished? If so, why, and how severely? If not, why not?

`Q6` [L16] You regularly purchase your games through a game distribution platform, which maintains a profile that includes your name, address, and snippets of your credit history (your payment was declined a couple of times, so the company has an "opinion" of your ability to pay), as well as a wealth of information about your gaming preferences and habits (do you really play THAT much?). In order to make some extra money, the company decides to sell this identifying information (not with your credit card number, of course!). If you knew about the sale, would you object? Do you think that such sales should be allowed as described, or prohibited by law? Does it matter whether the company has told you in advance by providing a 20-page legal document describing how it manages (and may sell) your personal data? What if the company includes a "font fingerprint" (based on your games) that, together with the geographic locale from your IP, enables web servers to uniquely identify you?

`Q7` [L16] Do you think that a person should be allowed to collect and sell information about other people, assuming that the person collecting the information is doing so in an acceptable, public way (not peeking into windows, for example)? In your answer, compare such a person with someone who makes their living that way (a tabloid photographer), but targets only a small number of people.

`Q8` [L17] Explain why the function $Q(x) = (222x + 777) \mod 256$ 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 we discussed in class.

`Q9` [L17] You download a copy of a new 3D editing tool from a website. You notice that the site, which is not the original author of the tool, also provides the SHA512 hash of the download. Explain the potential problem with using the hash provided from the website.

`Q10` [L17] A website publishes a collection of personal, encrypted documents along with the public key of a well-known media personality. Always a fan of the person, you download the documents and use the key to decrypt them. Unfortunately, the contents of the documents upset you, and you find it hard to believe that they were written by the person. But only they could have the private key! Right? Explain what else might have happened.

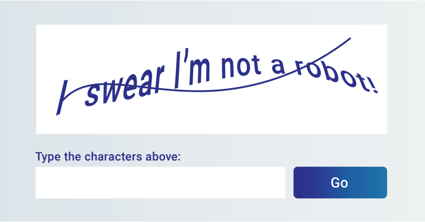
`Q11` [L17] While traveling in Europe, you spend a few nights in an inexpensive youth hostel, where you are told that you must use a proxy to access the Internet. All of your packets must be sent to the hostel's machine, which forwards them into the Internet. When you try to log in to machine X at UIUC to turn in a homework, however, you are told that the machine's key has changed. Do you accept the new key? Explain how the proxy might trick you into allowing it to see your communication with X, or explain why doing so is impossible.

`Q12a` [L13] You are a data scientist at HelloChef, an international company that delivers meal kits every week. Every weekly meal kit contains 4 meals, out of a selection of hundreds of meals that your kitchen can prepare. You are designing a recommendation engine to help your users pick the best meals for them. The recommendation engine uses collaborative filtering, i.e. it finds similar customers, and recommends meals that those similar customers have frequently ordered. Name some appropriate dimensions in this feature space of customers, that would help you find similarity between customers.

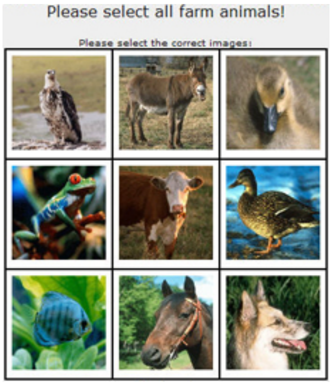
`Q12b` [L13] Now, as the data scientist at HelloChef, you have decided to build another recommendation engine. This time, you want to you content-based filtering, i.e. you want to find and recommend meals similar to what a particular customer has ordered in the past. Name some appropriate dimensions in this feature space of meals, that would help you find similarity between meals.

`Q13` [L11] A social media website wants to reduce spam on their website by blocking the use of bots. To differentiate between bots and humans, the website decides to ask for a CAPTCHA test every time you login or post frequently. Explain why each of these CAPTCHA tests are good or bad for this purpose:

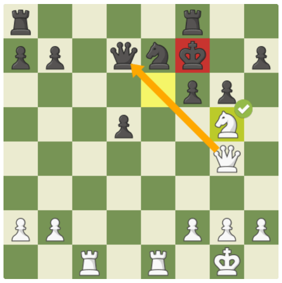
1. Text Captcha



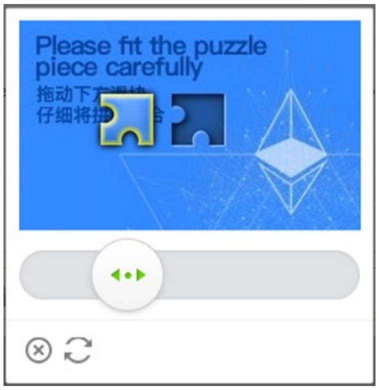
2. Selecting the correct group of images


Please select all farm animals!
Please select the correct images!

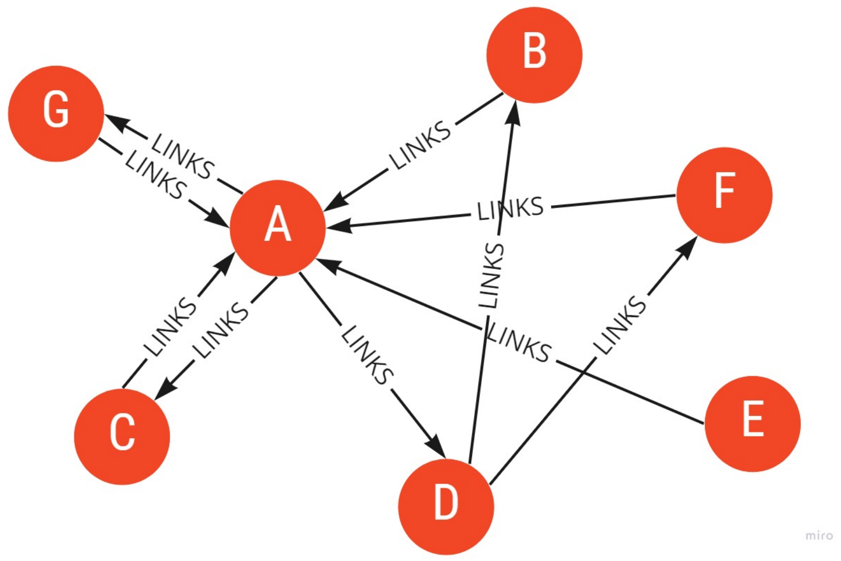
3. Winning a game of Chess



4. Sliding puzzle



`Q14` [L12] The following image shows a graph of websites and the links between them. Which website should have the highest PageRank, and why?



`Q15` [L14] When using Linear Regression to find a best fit line for our data points, what are we trying to minimize?

`Q16` [L15] Say we had data about the location of every student on campus throughout the day. By using Machine Learning with this location data as input, we can get several difference useful outputs. Give three examples of such useful outputs.

Questions with Solutions

`Q1` [L16] Explain the process of learning by example, and describe a situation in which a model learned by example might not reflect reality.

> In learning by example, one comes up with a rule that works for a set of examples—each example combines a specific sample with an answer, and the rule translates a sample into an answer. The rule is then hopefully accurate for other samples—those never before seen. For example, a young child might only have direct exposure to cats and dogs, and might "learn" that all animals are furry. The "rule" would then not be appropriate for fish, snakes, birds, and so forth.

`Q2` [L16] Heavy farm equipment is a rare sight on most highways in the North Pacific Coast of the US (Washington, Oregon, Northern California). Explain why autonomous vehicles sold in these regions must still be trained with simulations including such equipment.

> Rare events still occur, so autonomous vehicles operating in that region should be able to identify farm equipment. Also, people may travel, in which case the assumption of rarity may no longer be valid. Finally, the vehicle may be sold to someone who lives in a region in which farm equipment is common. Selling a vehicle that fails to operate properly in any of these scenarios is hazardous to both the owners and to the public.

`Q3` [L16] A company wants to provide a service to locate humans who might potentially be in danger from forest fires—whether because they have not followed evacuation orders, were hiking, or some other reason. Give four reasons that using ML for such a task is superior to hiring humans.

> Many of the reasons discussed in class might be appropriate. (1) Issues of human safety are particularly relevant here, though: the fewer humans need risk their lives to find the other humans, the better (humans will probably still be needed to perform the rescue, but that's not the question). (2) ML-laden drones are more maneuverable, cheaper, and faster than humans (unless the humans are in helicopters, which are absurdly expensive in comparison) in flying over/around forested areas, (3) Drones are easier and cheaper to purchase, maintain, and replicate than are humans skilled in this particular task—on the same note, some humans would need to be on duty at all times, whereas drones can simply be put in storage and pulled out when needed, and (4) drones can operate 24/7 (except for occasional recharging of batteries), and (5) can more easily leverage cameras/sensors based on non-visual frequency ranges when appropriate. (Any four of these five, or other reasonable answers, are fine.)

`Q4` [L16] During a heavy rain, an autonomous vehicle approaches an intersection in which a vandal has cut down the stop sign with a hacksaw. Open-source maps mark the intersection as a 4-way stop. Unfortunately, the vehicle fails to stop and hits another vehicle. Who do you think should be held responsible for the accident? Do you think a human driver would have avoided the accident? Explain your answers.

> Not clear that there's a "right" answer—humans who drive cautiously may avoid the accident (arguably), while those who are less cautious probably would not. Certainly humans are unlikely to make use of map data unless there's an app open that identifies the hazard and beeps loudly if the human isn't slowing down to stop, but that's essentially a sidekick AI, so not really the intention of the question. Open-source data are also of questionable value, since no one guarantees their accuracy, and malicious entities might contribute incorrect data. That all said, perhaps the autonomous system should be designed to drive cautiously, but already society has started complaining about such things, notably in the case of Teslas slowing down or braking "inexplicably." Humans do (fairly regularly) rear-end vehicles when a vehicle slows down quickly for a reason unknown to the human, so being cautious and conservative is not a panacea, either.

`Q5` [L16] The New York City Public Library allows citizens of the city to borrow physical library materials. Recently, the library acquired a set of e-books, which it lends electronically. Your friend, an Urbana resident from Hawaii, has managed to obtain a library card by using a VPN to trick the library into believing that they live in Manhattan. Assuming that your friend's deceit is detected, should they be punished? If so, why, and how severely? If not, why not?

> Again, one can argue either way: should the library be more careful and/or require physical presence instead of using an IP address? In reality, they probably are. But the friend's intention is clearly fraudulent, and exposes the library to potential theft (retrieving materials from outside the city and/or state is unrealistic). Is it against the law? Unclear.

`Q6` [L16] You regularly purchase your games through a game distribution platform, which maintains a profile that includes your name, address, and snippets of your credit history (your payment was declined a couple of times, so the company has an "opinion" of your ability to pay), as well as a wealth of information about your gaming preferences and habits (do you really play THAT much?). In order to make some extra money, the company decides to sell this identifying information (not with your credit card number, of course!). If you knew about the sale, would you object? Do you think that such sales should be allowed as described, or prohibited by law? Does it matter whether the company has told you in advance by providing a 20-page legal document describing how it manages (and may sell) your personal data? What if the company includes a "font fingerprint" (based on your games) that, together with the geographic locale from your IP, enables web servers to uniquely identify you?

> Same type of question—no "right" answer—just want you to think and respond. On the exam, we will be a little more specific to restrict your answers a bit. Most companies now do provide you with long documents about their plans for your information, along with the caveat that they may change the agreement at any time (so it's not really an agreement so much as just being able to say, "we told you so!").

`Q7` [L16] Do you think that a person should be allowed to collect and sell information about other people, assuming that the person collecting the information is doing so in an acceptable, public way (not peeking into windows, for example)? In your answer, compare such a person with someone who makes their living that way (a tabloid photographer), but targets only a small number of people.

> Many of the questions about other people gathering information on you are going to run up against examples that have been around for decades or longer. Certainly it will be hard to prohibit reporting of any activity classified as criminal (such as driving over the speed limit), for example.

`Q8` [L17] Explain why the function $Q(x) = (222x + 777) \mod 256$ 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 we discussed in class.

> In this case, the product (222 x) is always even, so the value of Q(x) is always odd. The smallest bit is thus not random at all.

`Q9` [L17] You download a copy of a new 3D editing tool from a website. You notice that the site, which is not the original author of the tool, also provides the SHA512 hash of the download. Explain the potential problem with using the hash provided from the website.

> Let's imagine that the website is provided by a malicious agent. The agent obtains a copy of the tool, adds a computer virus, and then posts the infected version to the web site. The SHA512 hash on the website is then computed based on the infected version. It does not match the hash on the original author's website, but if you don't check, you don't realize that fact.

`Q10` [L17] A website publishes a collection of personal, encrypted documents along with the public key of a well-known media personality. Always a fan of the person, you download the documents and use the key to decrypt them. Unfortunately, the contents of the documents upset you, and you find it hard to believe that they were written by the person. But only they could have the private key! Right? Explain what else might have happened.

> The website may be tricking you in regard to the public key's owner. Anyone can create a public-private key pair and attach any name to it. Without some better way of validating the authenticity of the public key's owner, one should not put any weight on the identity of the documents' author.

`Q11` [L17] While traveling in Europe, you spend a few nights in an inexpensive youth hostel, where you are told that you must use a proxy to access the Internet. All of your packets must be sent to the hostel's machine, which forwards them into the Internet. When you try to log in to machine X at UIUC to turn in a homework, however, you are told that the machine's key has changed. Do you accept the new key? Explain how the proxy might trick you into allowing it to see your communication with X, or explain why doing so is impossible.

> The proxy is attempting what is called a man/person-in-the-middle attack. The proxy gives you a new key. You encrypt your data using the new key and send it to the proxy. Instead of just forwarding the data, the proxy decrypts it (using the new key), reads it (collects data on you), and re-encrypts it with the UIUC machine's real key. The proxy also lies to the UIUC machine about your key, allowing it to receive responses meant for you, decrypt them, read them, and re-encrypt them using the new key so that you think your channel is secure.

`Q12a` [L13] You are a data scientist at HelloChef, an international company that delivers meal kits every week. Every weekly meal kit contains 4 meals, out of a selection of hundreds of meals that your kitchen can prepare. You are designing a recommendation engine to help your users pick the best meals for them. The recommendation engine uses collaborative filtering, i.e. it finds similar customers, and recommends meals that those similar customers have frequently ordered. Name some appropriate dimensions in this feature space of customers, that would help you find similarity between customers.

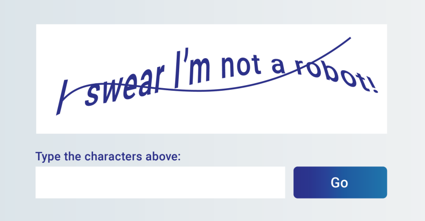
> Some useful features/dimensions of this feature space are:
>
> - Age
> - Dietary Preferences:
>   - Does the customer eat meat?
>   - Is the customer lactose intolerant?
> - Spice Tolerance
> - Location/Country
> - Average price of meals ordered

`Q12b` [L13] Now, as the data scientist at HelloChef, you have decided to build another recommendation engine. This time, you want to you content-based filtering, i.e. you want to find and recommend meals similar to what a particular customer has ordered in the past. Name some appropriate dimensions in this feature space of meals, that would help you find similarity between meals.

> Some useful features/dimensions of this feature space are:
>
> - Cuisine that the meal belongs to
> - Time required to prepare the meal
> - Cost
> - Spiciness
> - Does the meal contain meat?
> - Nutritional value

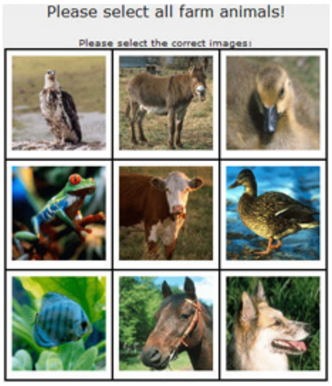
`Q13` [L11] A social media website wants to reduce spam on their website by blocking the use of bots. To differentiate between bots and humans, the website decides to ask for a CAPTCHA test every time you login or post frequently. Explain why each of these CAPTCHA tests are good or bad for this purpose:

1. Text Captcha



> Text CAPTCHAs used to be good. However, they are bad now, because text recognition has become trivial for computers now.
>
> *Note: you may have a different answer, which will be correct with reasonable explanation.*

2. Selecting the correct group of images

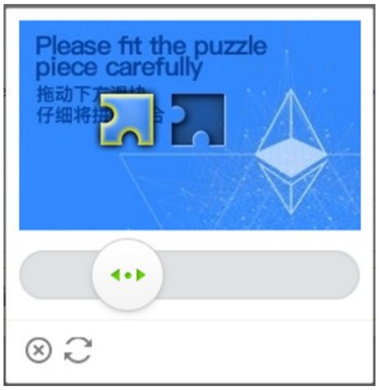

Please select all farm animals!
Please select the correct images:

> This is a good CAPTCHA test (for now). This is an easy and quick test for humans, and computers are just beginning to get good at understanding images. Moreover, this test requires additional understanding (such as what a "farm animal" is), which humans are good at.
>
> *Note: you may have a different answer, which will be correct with reasonable explanation.*

3. Winning a game of Chess

> This is a bad CAPTCHA test. Not all humans will be good at chess, and computers can easily beat humans in this game. Moreover, this is time consuming and not a quick test.

4. Sliding puzzle



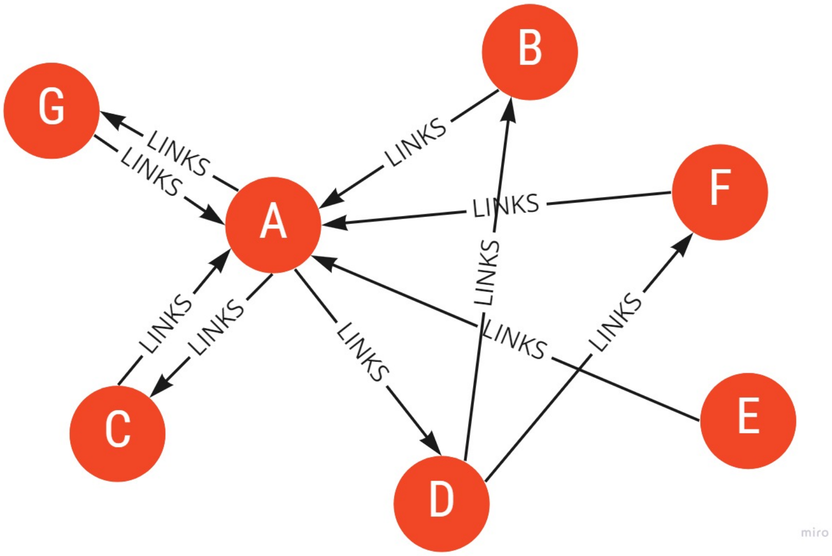
Please fit the puzzle piece carefully

拖动下方滑块

仔细将拼 合

> This is a good CAPTCHA test, as it is quick and easy for humans. It is not trivial for a computer to understand the puzzle.
>
> *Note: you may have a different answer, which will be correct with reasonable explanation.*

`Q14` [L12] The following image shows a graph of websites and the links between them. Which website should have the highest PageRank, and why?



> Website A would have the highest PageRank, because it has a large number of links pointing to it.

`Q15` [L14] When using Linear Regression to find a best fit line for our data points, what are we trying to minimize?

> The best fit line minimizes the sum of distances from all points to the line.

`Q16` [L15] Say we had data about the location of every student on campus throughout the day. By using Machine Learning with this location data as input, we can get several difference useful outputs. Give three examples of such useful outputs.

> Some useful outputs using Machine Learning are:
>
> - The optimal routes/spots for campus police to patrol, to minimize crime
> - The optimal location of bus stops that minimizes the time students spend getting from one class to another
> - The best places to install billboards for campus-wide announcements and advertising