Jeremy Tandjung - Assignment 1 CSS 490

**1) [5 points]** Give a **concrete example** of Data Object in the real word. Think about the attributes/features and **list all** attributes you can think of. A nice draw of a simple table will be awesome.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Observation** | **Age** | **Credit score** | **Annual income** | **Granted Loan** |
| Person1 | 23 | 756 | 60,000 | Yes |
| Person2 | 39 | 650 | 0 | No |
| Person3 | 40 | 700 | 120,000 | Yes |
| Person4 | 65 | 700 | 36,000 | Yes |
| Person5 | 30 | 687 | 40,000 | No |

**2) [5 points]** Briefly (1-2 sentences) justify your answers to the below questions.

1. You have developed an algorithm to automatically and efficiently retrieve addresses and phone numbers from a very large database, using a person’s name as the search key. Is this machine learning? Why or why not?

No, it is not an application of machine learning. Machine learning is a process that involves analyzing input to get an output that didn’t existed before.

In this case the algorithm did not output something new, it output something we already knew as each input (key) is uniquely associated with a value (addresses and phone numbers).

1. You have developed an algorithm to automatically and efficiently retrieve all similar images (could be ranked by similarity scores or probabilities) to a query image from a very large image database. Is this machine learning? Why or why not?

Yes, this is one of the applications of machine learning because we are analyzing input (the image we are comparing) and using that analysis to retrieve something new (similar images) as an output.

**3) [10 points, 2.5 each]** For each of the following scenarios, state which type of machine learning would be appropriate: classification, regression, clustering, or association rule analysis. Briefly (1-2 sentences) justify your answers.

1. You are a manager of a local Safeway store, and sales data indicates that if a customer buys onions and potatoes together, they are likely to also buy hamburger meat. You want to make more benefit from this by using promotional pricing or product placements.

In this scenario, an association rule analysis would be the appropriate technique as the manager wants to get information of the tendency of an item getting purchase by someone based on his or her previous purchase.

1. A dietician has been trying to understand how people’s dietary choices affect the amount of weight they gain or lose, but isn’t seeing obvious patterns. For a recent 6-month period, he has good records for 150 of his clients on their consumption of 12 different foods, along with the change in their weight over that period.

For this scenario, a regression algorithm would be appropriate because the dietician wants to know how does one’s diet could affect one’s diet and also which food affect the weight loss more than the other. Besides that, a regression would fit this scenario because we are looking for a trend.

1. You are having an argument with your friend about how many social groups there are at your school. You believe there are about half-a-dozen natural groups based on tastes in things like music, clothes, athletics, and politics, while your friend thinks everyone’s tastes are pretty random. You discover you can access (publicly available) individual records from a poll where 1000 students scored their preferences on 20 forms of arts and entertainment.

Since we are dealing with unknown data and looking for a pattern, a clustering algorithm would be appropriate for this scenario. The algorithm will cluster or “group” people based on their taste on a graph. However, if the datapoints are scattered around the graph then the friend’s opinion of everyone’s taste are fairly random.

1. You work at an oil company, and they are interested in predicting whether wells drilled in several new formations will produce oil or not. They give you a large quantity of data from past drilling efforts (geographic location, depth of well, type of rock, age of formation, etc.), along with the success or failure of each drilled well.

For this scenario, I would use a classification algorithm, specifically a binomial classification where the class is just producing oil or not based on the features given from the past dataset.

**4) [4 points]** Describe a situation in your life where machine learning might be of benefit. For example, this could be something you deal with at work, at school, or on the internet (e.g. a social networking site). Say as much as you can about the problem to be solved, the data or information you might collect, and the type of machine learning you think is applicable. (There isn’t necessarily a correct answer to this question; I just want you to start being able to recognize opportunities to apply machine learning.)

One of the applications of machine learning that I’m most excited of is applying machine learning algorithms to detect objects in an image or video. This area of study, which is called computer vision, could push humanity towards autonomous technology. We are now seeing many autonomous vehicle prototypes being tested by big tech companies like Tesla.

Besides that, object detection also could push technology to understand what is going on in an image or movie. Therefore, it could write stories based on an image or video and also make an image or video based on text. This will propel the creative industry to the next level.

**5) [6 points, 2 each]** Term project proposal

1. Which dataset and problem are you planning to work on for the term project? Why you are interested in this problem?

I’m doing a playing card detection project for the term project. The dataset I’m using can be found in Kaggle and it comprises of 292 photos of playing cards (multiple and single) in various condition and background.

I’m interested in this topic because I want to use this as practice to learn computer vision algorithms and making computer vision applications.

1. What kind of useful information/knowledge/pattern do you expect to discover from that dataset?

Color, x and y position.

1. Why you think your work will be valuable? (What is the potential future application of your work, or who will be your potential customers?)

If succeeded, this project could be applied to assistive technology for the disabled people. Besides that, it could also replace professional card dealers in casinos.