CISC 3440: Homework 1

Due 9/11/23 at 11:59 PM

15 points

1. (5 points) Describe a problem other than traffic lights (lecture example) and spam filtering (book example) that is hard to solve with traditional programming techniques, but could be easier to solve with machine learning. Please include the following in your answer:

    a) Describe why the problem is so hard to solve with traditional programming techniques. Identify a few specific issues.

    b) Assuming a supervised approach, what kind of data would you need to perform machine learning? What inputs would the machine learning model accept as input, and what labels would it require? Where would the data come from?

2. (2 points) What is the difference between a labeled and unlabeled training set?

For the next two questions, refer to the life satisfaction score dataset:
https://github.com/ageron/data/raw/main/lifesat/lifesat.csv

3. (3 points) In class, we estimated different countries' life satisfaction scores with a linear model. The model's parameters are as follows:

    $\theta_0 = 3.75$
    $\theta_1 = 6.78 \times 10^{-5}$

    Estimate the life satisfaction score for each of the following countries with their GDP as input, and indicate how close it is to their true value. Please show your work:

    a) United Kingdom
    b) Belgium
    c) Germany

4. (5 points) Consider the following alternative linear model for life satisfaction scores:

   $\theta_0 = 6.17$
   $\theta_1 = 1.52 \times 10^{-5}$

   Estimate the life satisfaction score for each of the following countries with their GDP as input, and indicate how close it is to their true value. Please show your work:

   a) United Kingdom
   b) Belgium
   c) Germany

   Is this model better or worse than the model in Question 3? Why?