

COGS9: Introduction to Data Science

Assignment #3: Machine Learning (20 points)

Submit through Gradescope individually

First Name	Last Name	PID

Once complete, download as a PDF and submit on Gradescope.

Instructions: For this assignment, you'll be using interactive widgets throughout this piece from the Parametric Press, titled "The Myth of the Impartial Machine". Navigate to the following URL to get started: <https://parametric.press/issue-01/the-myth-of-the-impartial-machine/>. Read and understand the whole piece first and then go through to answer the questions throughout. This article might also make more sense after our first machine learning (ML) talk, if you are not already familiar with ML. When you are done **download as a PDF and submit on Gradescope.**

Question 1. (2 pt) In a few sentences, what is the relationship between machine learning and artificial intelligence?

REPLACE THIS TEXT WITH YOUR ANSWER.

Question 2. (3 pts) Sampling Bias: We often aren't able to collect information about entire populations and are forced to take a sample from the population. In the "Sampling Errors Can Lead to Biased Models" Interactive widget, collect 10 samples, each of sample size 10. Record the sample mean for each sample you collect in the table:

Sample	1	2	3	4	5	6	7	8	9	10
Sample Size										
Sample mean										

(i.) Interpret the results of your 10 samples, commenting on the effects of sampling bias and how it relates to the true population mean.

(ii.) In a few sentences, compare and contrast sampling errors and non-sampling errors.

REPLACE THIS TEXT WITH YOUR INTERPRETATION.

Question 3. (3 pts) Briefly explain the sampling bias issue underlying the ImageNet data discussed in the article. If you were in charge, brainstorm at least one way you would go about addressing this bias.

REPLACE THIS TEXT WITH YOUR ANSWER.

Question 4 (3 pts). In the “Algorithms can amplify bias found in data” section, adjust the bias and accuracy sliders for five different combinations. Record each combination you try in the table below.

	Bias	Model Accuracy	Result
ex	70/30	70%	<i>Not incentivized to amplify bias</i>
1			
2			
3			
4			
5			

Using the data you collected, explain the relationship between model bias and accuracy to determine when models are incentivized to amplify bias.

REPLACE THIS TEXT WITH YOUR ANSWER.

Question 5. (3 pts) In the Feedback Loops Runaway simulation set the simulator to have the same number of crimes occur in precinct A and precinct B to be "4". Run the simulation, taking the time to understand the results as the simulation runs. Explain the results of the simulation, using the outcome of the simulation to explain what has happened with regards to crime and policing.

REPLACE THIS TEXT WITH YOUR EXPLANATION.

Question 6. (3 pts) Choose one of the three methods explained in the article for how people are tackling bias that you think is most important. Explain briefly (a few sentences) (i.) what approach you choose (from the article). Then, (ii.) think critically and explain the pros and cons of this approach. We are looking for critical thinking, not that one position is right or wrong.

REPLACE THIS TEXT WITH YOUR ANSWER.

Question 7. (3 pts) How do you envision the issues discussed in this piece (biased data, biased algorithms, and attempts to limit bias) changing over time. Will they improve? Stay the same? Worsen? Defend your position. There is no right or wrong answer, we are looking for critical thinking applied to the question.

REPLACE THIS TEXT WITH YOUR ANSWER.

Once complete, download as a PDF and submit on Gradescope.