

## 15-110 Check6-1 - Written Portion

**Name:**

**AndrewID:**

---

Complete the following problems in the fillable PDF, or print out the PDF, write your answers by hand, and scan the results.

When you are finished, upload your check6-1.pdf to **Check6-1 - Written** on Gradescope. Make sure to upload your Check6-1 work for the Hw6 project as well!

### Written Problems

[#1 - Data Formats - 5pts](#)

[#2 - Parsing Data - 11pts](#)

[#3 - Components vs. Rules - 5pts](#)

[#4 - Simulation Code - 9pts](#)

[#5 - Machine Learning Categories - 10pts](#)

[#6 - Machine Learning Process - 5pts](#)

# Written Problems

## #1 - Data Formats - 5pts

*Can attempt after Data Analysis I lecture*

For each of the following data formats, identify whether it would be best interpreted as a CSV file, a JSON file, or as plaintext.

```
{ "restaurants": [
  { "restaurant" : "Chipotle",
    "menu" :
      { "lunch" :
          { "burrito" : 7.99,
            "tacos" : 6.99,
            "bowl" : 8.99 },
        "dinner" :
          { "burrito" : 8.99,
            "tacos" : 7.99,
            "bowl" : 9.99 } } },
  { "restaurant" : "Sushi Fuku",
    "menu":
      { "lunch" :
          { "bowl" : 8.99,
            "sushi" : 10.99 },
        "dinner":
          { "bowl" : 10.99,
            "sushi" : 12.99 } } } ] }
```

- CSV
- JSON
- Plaintext

City,Longitude,Latitude

Los Angeles,34°03'N,118°15'W

New York City,40°42'46"N,74°00'21"W

Paris,48°51'24"N,2°21'03"E

- CSV
- JSON
- Plaintext

## #2 - Parsing Data - 11pts

*Can attempt after Data Analysis I lecture*

You have been given a set of data about CMU classes in the following format (whitespace added for clarity):

Professor,	ClassNum,	Days,	Time
Cortina,	15104,	MTWF,	09:05-09:55
Rivers,	15110,	MWThF,	14:30-15:20
Xhakaj,	15110,	MWThF,	15:35-16:25
Xhakaj,	15121,	TWTh,	11:50-13:10

Assume you've already split the string on "\n" and used the variable `row` to iterate through each class one line at a time.

How would you determine which department each class is in? Recall that the first two digits of the class number indicate the department.

- `row.split(',')[1][0:1]`
- `row.split(',')[1][0:2]`
- `row.split(',')[1][2:]`

How would you determine the start time and end time of a class and set those times as strings in the variables `start` and `end`? **Select all lines that are needed.** Assume that the code is run from the top selected line to the bottom selected line.

- `times = row.split(',')[0]`
- `times = row.split(',')[3]`
  
- `start = row.split('-')[0]`
- `start = times.index('-') - 1`
- `start = times.split('-')[0]`
  
- `end = times.index('-') + 1`
- `end = times.split('-')[1]`
- `end = start + "1:00"`

### #3 - Components vs. Rules - 5pts

*Can attempt after Simulation I lecture*

Let's say we want to design a simulation that determines how many students will sign up for a course during registration week. The simulation's time loop will loop over each sign-up time slot in order.

We need to design the model for this simulation. For each of the following values, would this value work better as a **component** of the model, or as a **rule** of the model?

Current length of the course's waitlist

- Component
- Rule

Students are more likely to sign up if a class is required for their major

- Component
- Rule

Number of students who are required to take this class, and haven't taken it yet, organized by sign-up timeslot

- Component
- Rule

Students are less likely to sign up for a class if the waitlist is long

- Component
- Rule

Information on whether or not the course will be offered again in the following semester

- Component
- Rule

## #4 - Simulation Code - 9pts

*Can attempt after Simulation I lecture*

We want to write code for a simulation that moves a circle from the left side of the screen to the right side of the screen in a 400px x 400px window.

For each part of the simulation (the Model, the View, and the Time Rules), select the line of code that needs to be included in that part.

Hint: if you're not sure, try implementing this using the simulation starter code!

Which line of code should be included in the **model**, in `makeModel(data)`?

- `x = 5`
- `data["left"] = 5`
- `canvas.create_oval(x, y, x + 40, y + 40)`

Which line of code should be included in the **view**, in `makeView(data, canvas)`?

- `data["left"] = data["left"] + 5`
- `canvas.create_oval(200, 200, 200 + 40, 200 + 40)`
- `canvas.create_oval(data["left"], 200, data["left"] + 40, 200 + 40)`

Which line of code should be included in the **time rules**, in `runRules(data, call)`?

- `data["left"] = 5`
- `data["left"] = data["left"] + 5`
- `x = data["left"] + 5`

## #5 - Machine Learning Categories - 10pts

*Can attempt after Machine Learning lecture*

For each of the following prompts, fill in the blanks with the type of **learning algorithm** that should be used and/or the type of **reasoning algorithm** that should be used.

You have a dataset that consists of student grades from past semesters of 15-110, including final grades. Use \_\_\_\_\_ learning to predict a student's **numerical** final grade based on their **numerical** quiz scores with a \_\_\_\_\_ algorithm.

You have a dataset of weather patterns in different major cities around the world. Use \_\_\_\_\_ learning to propose **new groupings** of cities based on the **categorical** weather patterns with a \_\_\_\_\_ algorithm.

You have a dataset of athletes' descriptions (age, height, weight, etc) and the sport that they play. Use \_\_\_\_\_ learning to predict an athlete's **categorical** sport based on their **numerical** age, height, weight, etc. with a \_\_\_\_\_ algorithm.

To train a robot how to throw a basketball through a hoop through repeated practice and feedback, you'd want to use \_\_\_\_\_ learning.

To identify previously-unknown market trends based on stocks that go up and down together at similar times, you'd want to use \_\_\_\_\_ learning.

## #6 - Machine Learning Process - 5pts

*Can attempt after Machine Learning lecture*

Imagine a scenario where Bill wants to train a machine learning algorithm to identify which pictures on the internet have cats in them. He downloads 1,000 pictures of cats and other animals from the internet, decides to use a basic image recognition algorithm which will identify important features, trains on all 1,000 pictures, then tests his on a quarter of that dataset (250 pictures). He finds that his algorithm has a 97% success rate, which he publishes on his blog.

Bill made a few mistakes in this process. **What was his biggest mistake?**