# XCS221 Assignment 7 — Logic (Extra Credit)

**Due Sunday, Jan. 29 at 11:59pm PT.**

### Guidelines

1. If you have a question about this homework, we encourage you to post your question on our Slack channel, at http://xcs221-scpd.slack.com/
2. Familiarize yourself with the collaboration and honor code policy before starting work.
3. For the coding problems, you must use the packages specified in the provided environment description. Since the autograder uses this environment, we will not be able to grade any submissions which import unexpected libraries.

### Submission Instructions

**Written Submission:** Some questions in this assignment require a written response. For these questions, you should submit a PDF with your solutions online in the online student portal. As long as the PDF is legible and organized, the course staff has no preference between a handwritten and a typeset LaTeX submission. If you wish to typeset your submission and are new to LaTeX, you can get started with the following:

- Type responses only in `submission.tex`.
- Submit the compiled PDF, **not** `submission.tex`.
- Use the commented instructions within the `Makefile` and `README.md` to get started.

Also note that your answers should be in order and clearly and correctly labeled to receive credit. Be sure to submit your final answers as a PDF and **tag all pages correctly when submitting to Gradescope**.

**Coding Submission:** Some questions in this assignment require a coding response. For these questions, you should submit only the `src/submission.py` file in the online student portal. For further details, see Writing Code and Running the Autograder below.

### Honor code

We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down the solutions independently, and without referring to written notes from the joint session. In other words, each student must understand the solution well enough in order to reconstruct it by him/herself. In addition, each student should write on the problem set the set of people with whom s/he collaborated. Further, because we occasionally reuse problem set questions from previous years, we expect students not to copy, refer to, or look at the solutions in preparing their answers. It is an honor code violation to intentionally refer to a previous year's solutions. More information regarding the Stanford honor code can be foudn at https://communitystandards.stanford.edu/policies-and-guidance/honor-code.

### Writing Code and Running the Autograder

All your code should be entered into `src/submission.py`. When editing `src/submission.py`, please only make changes between the lines containing `### START_CODE_HERE ###` and `### END_CODE_HERE ###`. Do not make changes to files other than `src/submission.py`.

The unit tests in `src/grader.py` (the autograder) will be used to verify a correct submission. Run the autograder locally using the following terminal command within the `src/` subdirectory:

```
$ python grader.py
```

There are two types of unit tests used by the autograder:

- `basic`: These tests are provided to make sure that your inputs and outputs are on the right track, and that the hidden evaluation tests will be able to execute.
- `hidden`: These unit tests are the evaluated elements of the assignment, and run your code with more complex inputs and corner cases. Just because your code passed the basic local tests does not necessarily mean that they will pass all of the hidden tests. These evaluative hidden tests will be run when you submit your code to the Gradescope autograder via the online student portal, and will provide feedback on how many points you have earned.

For debugging purposes, you can run a single unit test locally. For example, you can run the test case `3a-0-basic` using the following terminal command within the `src/` subdirectory:

```
$ python grader.py 3a-0-basic
```

Before beginning this course, please walk through the Anaconda Setup for XCS Courses to familiarize yourself with the coding environment. Use the env defined in `src/environment.yml` to run your code. This is the same environment used by the online autograder.

### Test Cases

The autograder is a thin wrapper over the python `unittest` framework. It can be run either locally (on your computer) or remotely (on SCPD servers). The following description demonstrates what test results will look like for both local and remote execution. For the sake of example, we will consider two generic tests: `1a-0-basic` and `1a-1-hidden`.

### Local Execution - Hidden Tests

All hidden tests rely on files that are not provided to students. Therefore, the tests can only be run remotely. When a hidden test like `1a-1-hidden` is executed locally, it will produce the following result:

```
----- START 1a-1-hidden:  Test multiple instances of the same word in a sentence.
----- END 1a-1-hidden [took 0:00:00.011989 (max allowed 1 seconds), ???/3 points] (hidden test ungraded)
```

### Local Execution - Basic Tests

When a basic test like `1a-0-basic` passes locally, the autograder will indicate success:

```
----- START 1a-0-basic:  Basic test case.
----- END 1a-0-basic [took 0:00:00.000062 (max allowed 1 seconds), 2/2 points]
```

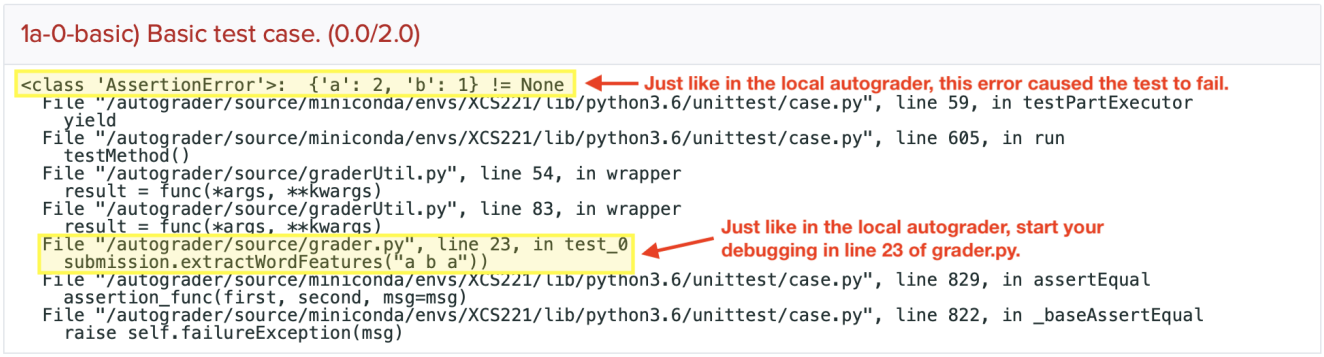
When a basic test like `1a-0-basic` fails locally, the error is printed to the terminal, along with a stack trace indicating where the error occurred:

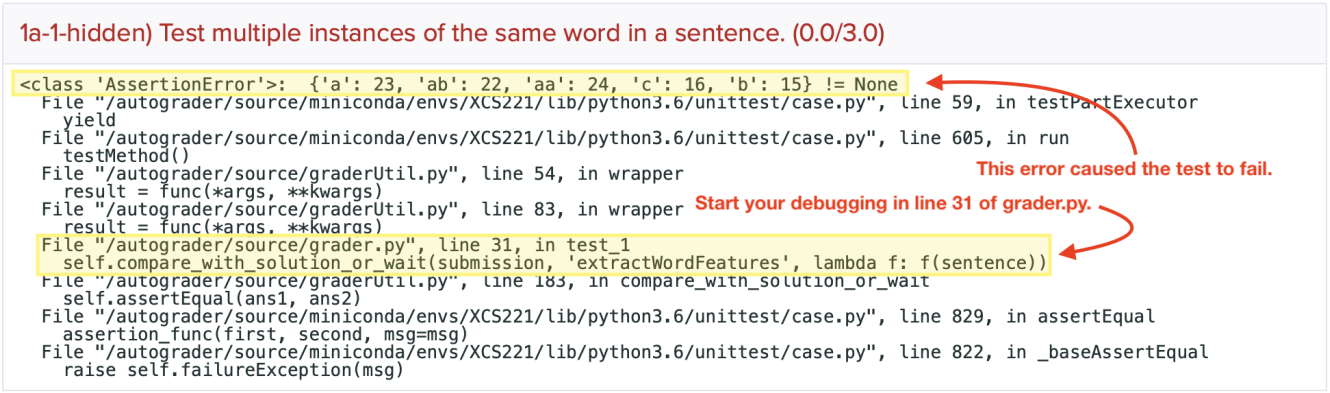
```
----- START 1a-0-basic:  Basic test case.
<class 'AssertionError'>
{'a': 2, 'b': 1} != None
  File "/Users/grinch/Local_Documents/Software/anaconda3/envs/XCS221/lib/python3.6/unittest/case.py", line 59, in testPartExecutor
    yield
  File "/Users/grinch/Local_Documents/Software/anaconda3/envs/XCS221/lib/python3.6/unittest/case.py", line 605, in run
    testMethod()
  File "/Users/grinch/Local_Documents/SCPD/XCS221/A1/src/graderUtil.py", line 54, in wrapper
    result = func(*args, **kwargs)
  File "/Users/grinch/Local_Documents/SCPD/XCS221/A1/src/graderUtil.py", line 83, in wrapper
    result = func(*args, **kwargs)
  File "/Users/grinch/Local_Documents/SCPD/XCS221/A1/src/grader.py", line 23, in test_0
    submission.extractWordFeatures("a b a"))
  File "/Users/grinch/Local_Documents/Software/anaconda3/envs/XCS221/lib/python3.6/unittest/case.py", line 829, in assertEqual
    assertion_func(first, second, msg=msg)
  File "/Users/grinch/Local_Documents/Software/anaconda3/envs/XCS221/lib/python3.6/unittest/case.py", line 822, in _baseAssertEqual
    raise self.failureException(msg)
----- END 1a-0-basic [took 0:00:00.003809 (max allowed 1 seconds), 0/2 points]
```



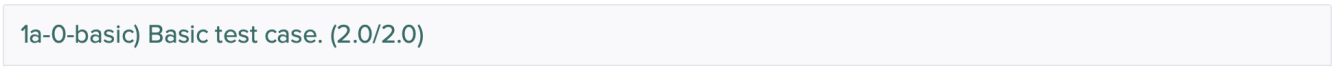
### Remote Execution

Basic and hidden tests are treated the same by the remote autograder. Here are screenshots of failed basic and hidden tests. Notice that the same information (error and stack trace) is provided as the in local autograder, now for both basic and hidden tests.




Finally, here is what it looks like when basic and hidden tests pass in the remote autograder.




**Introduction**

In this assignment, you will get some hands-on experience with logic. You'll see how logic can be used to represent the meaning of natural language sentences, and how it can be used to solve puzzles and prove theorems. Most of this assignment will be translating English into logical formulas, but in Problem 4, you will delve into the mechanics of logical inference.

To get started, launch a Python shell and try typing the following commands to add logical expressions into the knowledge base.

```
(XCS221) $ python
Python 3.6.9
Type "help", "copyright", "credits" or "license" for more information.
>>> from logic import *
>>> Rain = Atom('Rain')           # Shortcut
>>> Wet = Atom('Wet')             # Shortcut
>>> kb = createResolutionKB()     # Create the knowledge base
>>> kb.ask(Wet)
I don't know.
>>> kb.ask(Not(Wet))
I don't know.
>>> kb.tell(Implies(Rain, Wet))
I learned something.
>>> kb.ask(Wet)
I don't know.
>>> kb.tell(Rain)
I learned something.
>>> kb.tell(Wet)
I already knew that.
>>> kb.ask(Wet)
Yes.
>>> kb.ask(Not(Wet))
No.
>>> kb.tell(Not(Wet))
I don't buy that.
```

To print out the contents of the knowledge base, you can call `kb.dump()`. For the example above, you get:

```
==== Knowledge base [3 derivations] ===
* Or(Not(Rain),Wet)
* Rain
- Wet
```

In the output, '*' means the fact was explicitly added by the user, and '-' means that it was inferred.

Here is a table that describes how logical formulas are represented in code. Use it as a reference guide:

| **Name** | **Mathematical Notation** | **Code** |
|---|---|---|
| Constant symbol | stanford | `Constant('stanford')` (must be lowercase) |
| Variable symbol | $x$ | `Variable('$x')` (must be lowercase) |
| Atomic formula (atom) | Rain | `Atom('Rain')` (predicate must be uppercase) |
| | LocatedIn(stanford, $x$) | `Atom('LocatedIn','stanford','$x')` (arguments are symbols) |
| Negation | ¬Rain | `Not(Atom('Rain'))` |
| Conjunction | Rain ∧ Snow | `And(Atom('Rain'), Atom('Snow'))` |
| Disjunction | Rain ∨ Snow | `Or(Atom('Rain'), Atom('Snow'))` |
| Implication | Rain → Wet | `Implies(Atom('Rain'), Atom('Wet'))` |
| Equivalence | Rain ↔ Wet (syntactic sugar for: Rain → Wet ∧ Wet → Rain) | `Equiv(Atom('Rain'), Atom('Wet'))` |
| Existential quantification | $\exists x$.LocatedIn(stanford, $x$) | `Exists('$x', Atom('LocatedIn', 'stanford', '$x'))` |
| Universal quantification | $\forall x$.MadeOfAtoms($x$) | `Forall('$x', Atom('MadeOfAtoms', '$x'))` |

The operations `And` and `Or` only take two arguments. If we want to take a conjunction or disjunction of more than two, use `AndList` and `OrList`. For example: `AndList([Atom('A'), Atom('B'), Atom('C')])` is equivalent to `And(And(Atom('A'), Atom('B')), Atom('C'))`.

1. **Propositional Logic**

Write a propositional logic formula for each of the following English sentences in the given function in `submission.py`. For example, if the sentence is *"If it is raining, it is wet"*, then you would write `Implies(Atom('Rain'), Atom('Wet'))`, which would be Rain $\rightarrow$ Wet in symbols (see `examples.py`).

*Note: Don't forget to return the constructed formula!*

(a) **[1 point (Coding, Extra Credit)]** *"If it's summer and we're in California, then it doesn't rain."*

(b) **[1 point (Coding, Extra Credit)]** *"It's wet if and only if it is raining or the sprinklers are on."*

(c) **[2 points (Coding, Extra Credit)]** *"Either it's day or night (but not both)."*

You can run the following command to test each formula:

```
python grader.py 1a-0-basic
```

If your formula is wrong, then the grader will provide a counterexample, which is a model that your formula and the correct formula don't agree on. For example, if you accidentally wrote `And(Atom('Rain'), Atom('Wet'))` for *"If it is raining, it is wet"*, then the grader would output the following:

```
Your formula (And(Rain,Wet)) says the following model is FALSE, but it should be TRUE:
* Rain = False
* Wet = True
* (other atoms if any) = False
```

In this model, it is not raining and it is wet, which satisfies the correct formula Rain $\rightarrow$ Wet (`TRUE`), but does not satisfy the incorrect formula Rain $\wedge$ Wet (`FALSE`). Use these counterexamples to guide you in the rest of the assignment.

2. **First Order Logic**

Write a first-order logic formula for each of the following English sentences in the given function in `submission.py`. For example, if the sentence is *"There is a light that shines"*, then you would write `Exists('$x', And(Atom('Light', '$x'), Atom('Shines', '$x')))`, which would be $\exists x.\text{Light}(x) \wedge \text{Shines}(x)$ in symbols (see `examples.py`).

*Tip: Python tuples can span multiple lines, which help with readability when you are writing logic expressions (some of them in this homework can get quite large)*

(a) **[0.50 points (Coding, Extra Credit)]** *"Every person has a mother."*

(b) **[0.50 points (Coding, Extra Credit)]** *"At least one person has no children."*

(c) **[0.50 points (Coding, Extra Credit)]** Create a formula which defines `Daughter(x,y)` in terms of `Female(x)` and `Child(x,y)`.

(d) **[0.50 points (Coding, Extra Credit)]** Create a formula which defines `Grandmother(x,y)` in terms of `Female(x)` and `Parent(x,y)`.

3. **Liar Puzzle**

Someone crashed the server, and accusations are flying. For this problem, we will encode the evidence in first-order logic formulas to find out who crashed the server. You've narrowed it down to four suspects: John, Susan, Mark, and Nicole. You have the following information:

- John says: "It wasn't me!"
- Susan says: "It was Nicole!"
- Mark says: "No, it was Susan!"
- Nicole says: "Susan's a liar."
- You know that exactly one person is telling the truth.
- You also know exactly one person crashed the server.

(a) **[2 points (Coding, Extra Credit)]** Fill out `liar()` to return a list of 6 formulas, one for each of the above facts. Be sure your formulas are exactly in the order specified.

You can test your code using the following commands:

```
python grader.py 3a-0-basic
python grader.py 3a-1-basic
python grader.py 3a-2-basic
python grader.py 3a-3-basic
python grader.py 3a-4-basic
python grader.py 3a-5-basic
python grader.py 3a-all-basic  # Tests the conjunction of all the formulas
```

To solve the puzzle and find the answer, `tell` the formulas to the knowledge base and `ask` the query `CrashedServer('$x')`, by running:

```
python grader.py 3a-run-basic
```

4. **Odd and Even Integers**

In this problem, we will see how to use logic to automatically prove mathematical theorems. We will focus on encoding the theorem and leave the proving part to the logical inference algorithm. Here is the theorem:

If the following constraints hold:

1. Each number $x$ has exactly one successor, which is not equal to $x$.
2. Each number is either odd or even, but not both.
3. The successor of an even number is odd.
4. The successor of an odd number is even.
5. For every number $x$, the successor of $x$ is larger than $x$.
6. Larger is a transitive property: if $x$ is larger than $y$ and $y$ is larger than $z$, then $x$ is larger than $z$.

Then we have the following consequence:

- For each number, there is an even number larger than it.

*Note: in this problem, "larger than" is just an arbitrary relation, and you should not assume it has any prior meaning. In other words, don't assume things like "a number can't be larger than itself" unless explicitly stated.*

(a) **[2 points (Coding, Extra Credit)]** Fill out `ints()` to construct six formulas for each of the constraints. The consequence has been filled out for you (`query` in the code). You can test your code using the following commands:

```
python grader.py 4a-0-basic
python grader.py 4a-1-basic
python grader.py 4a-2-basic
python grader.py 4a-3-basic
python grader.py 4a-4-basic
python grader.py 4a-5-basic
python grader.py 4a-all-basic  # Tests the conjunction of all the formulas
```

To finally prove the theorem, `tell` the formulas to the knowledge base and `ask` the query by running model checking (on a finite model):

```
python grader.py 4a-all-basic
```

5. **Semantic Parsing**

Semantic parsing is the task of converting natural lanugage utterances into first-order logic formulas. We have created a small set of grammar rules in the code for you in `createBaseEnglishGrammar()`. In this problem, you will add additional grammar rules to handle a wider variety of sentences. Specifically, create a `GrammarRule` for each of the following sentence structures.

(a) **[2 points (Coding, Extra Credit)]** Example: *Every person likes some cat.*

General template:

```
$Clause ← every $Noun $Verb some $Noun
```

(b) **[2 points (Coding, Extra Credit)]** Example: *There is some cat that every person likes.*

General template:

```
$Clause ← there is some $Noun that every $Noun $Verb
```

(c) **[2 points (Coding, Extra Credit)]** Example: *If a person likes a cat then the former feeds the latter.*

General template:

```
$Clause ← if a $Noun $Verb a $Noun then the former $Verb the latter
```

After implementing these functions, you should be able to try some simple queries using `nli.py`! For example:

```
(XCS221) $ python nli.py
============================================================
Hello!  Talk to me in English.
Tell me something new (end the sentence with '.') or ask me a question (end the sentence with '?')
    .
Type 'help' for additional commands.
------------------------------
> Every person likes some cat.

>>>>> I learned something.
------------------------------
> Every cat is a mammal.

>>>>> I learned something.
------------------------------
> Every person likes some mammal?

>>>>> Yes.
```

6. **Ethical Issue Spotting**

One of the goals of this course is to teach you how to tackle real-world problems with tools from AI. But real-world problems have real-world consequences. Along with technical skills, an important skill every practitioner of AI needs to develop is an awareness of the ethical issues associated with AI. The purpose of this exercise is to practice spotting potential ethical concerns in applications of AI - even seemingly innocuous ones.

In this question, you will explore the ethics of four different real-world scenarios using the ethics guidelines produced by a machine learning research venue, the NeurIPS conference. The NeurIPS Ethical Guidelines list sixteen non-exhaustive concerns under Potential Negative Social Impacts and General Ethical Conduct (the numbered lists). For each scenario, you will write a potential negative impacts statement. To do so, you will first determine if the algorithm / dataset / technique could have a potential negative social impact or violate general ethical conduct (again, the sixteen numbered items taken from the NeurIPS Ethical Guidelines page). If the scenario does violate ethical conduct or has potential negative social impacts, list one concern it violates and justify why you think that concern applies to the scenario. If you do **not** think the scenario has an ethical concern, explain how you came to that decision. Unlike earlier problems in the homework there are many possible good answers. If you can justify your answer, then you should feel confident that you have answered the question well.

Each of the scenarios is drawn from a real AI research paper. The ethics of AI research closely mirror the potential real-world consequences of deploying AI, and the lessons you'll draw from this exercise will certainly be applicable to deploying AI at scale. As a note, you are **not** required to read the original papers, but we have linked to them in case they might be useful. Furthermore, you are welcome to respond to anything in the linked article that's not mentioned in the written scenario, but the scenarios as described here should provide enough detail to find at least one concern.

**What we expect:** A 2-5 sentence paragraph for each of the scenarios where you either A. identify at least one ethical concern from the NeurIPS Ethical Guidelines and justify why you think it applies, or B. state that you don't think a concern exists and justify why that's the case. Chosen scenarios may have anywhere from zero to multiple concerns that match, but you are only required to pick one concern (if it exists) and justify your decision accordingly. Furthermore, copy out and underline the ethical checklist item to which you are referring as part of your answer (i.e.: Severely damage the environment). We have also included a citation in the example solution below, but you are not required to add citations to your response.

**Example Scenario:** You work for a U.S. hospital that has recently implemented a new intervention program that enrolls at-risk patients in programs to help address their chronic medical issues proactively before the patients end up in the hospital. The intervention program automatically identifies at-risk patients by predicting patients' risk scores, which are measured in terms of healthcare costs. However, you notice that for a given risk score tier, the Black patients are considerably sicker when enrolled than white patients, even though their assigned illness risk score is identical. You manually re-assign patients' risk scores based on their current symptoms and notice that the percentage of Black patients who would be enrolled has increased from 17% to over 45% [1].

**Example Solution:** This algorithm has likely encoded, contains, or potentially exacerbates bias against people of a certain race or ethnicity since the algorithm predicts healthcare costs. Because access to medical care in the U.S. is unequal, Black patients tend to have lower healthcare costs than their white counterparts [2]. Thus the algorithm will incorrectly predict that they are at lower risk.

(a) **[1 point (Written, Extra Credit)]**

An investment firm develops a simple machine learning model to predict whether an individual is likely to default on a loan from a variety of factors, including location, age, credit score, and public record. After looking through their results, you find that the model predicts mainly based on location and that the model

[1] Obermeyer et al. Dissecting racial bias in an algorithm used to manage the health of populations. 2019.

[2] Institue of Medicine of the National Academies. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. 2003

mainly accepts loans from urban centers and denies loans from rural applicants [3]. Furthermore, looking at the gender and ethnicity of the applicants, you find that the model has a significantly higher false positive rate for Black and male applicants than for other groups. In a false positive prediction, a model misclassifies someone who does not default as likely to default.

(b) **[1 point (Written, Extra Credit)]**

Stylometry is a way of predicting the author of contested or anonymous text by analyzing the writing patterns in the anonymous text and other texts written by the potential authors. Recently, highly accurate machine learning algorithms have been developed for this task. While these models are typically used to analyze historical documents and literature, they could be used for deanonymizing a wide range of texts, including code [4].

(c) **[1 point (Written, Extra Credit)]**

A research group scraped millions of faces of celebrities off of Google images to develop facial recognition technology [5]. The celebrities did not give permission for their images to be used in the dataset and many of the images are copyrighted. For copyrighted photos, the dataset provides URL links to the original image along with bounding boxes for the face.

(d) **[1 point (Written, Extra Credit)]**

Researchers have recently created a machine learning model that can predict plant species automatically directly from a single photo [6]. The model was trained using photos uploaded to the iNaturalist app by users who consented to use of their photos for research purposes, and the model is only used within the app to help users identify plants they might come across in the wild.

[3] Imperial College London. Loan Default Prediction Dataset. 2014.
[4] Caliskan-Islam et. al. De-anonymizing programmers via code stylometry. 2015.
[5] Parkhi et al. VGG Face Dataset. 2015.
[6] iNaturalist. A new vision model. 2020.

This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset LaTeX solutions.

---

# 6.a

# 6.b

## 6.c

# 6.d