

## P2# Programming with python

### Problem 1:

In this problem, you will implement a bunch of short functions. The main purpose of this exercise is to familiarize yourself with Python, but as a bonus, the functions that you will implement will come in handy in subsequent homeworks.

**Do not import any outside libraries (e.g. numpy).** Only standard Python libraries and/or the libraries imported in the starter code are allowed.

If you're new to Python, the following provide pointers to various tutorials and examples for the language:

- [Python for Programmers](https://wiki.python.org/moin/BeginnersGuide/Programmers)  
(<https://wiki.python.org/moin/BeginnersGuide/Programmers> )
- [Example programs of increasing complexity](https://wiki.python.org/moin/SimplePrograms)  
(<https://wiki.python.org/moin/SimplePrograms> )

**What we expected:** Python code implementing the functions provided in submission.py. Try to make your code as clean and simple as possible and be sure to write your answers between the begin answer and end answer comments.

- [3points] Implement `find_alphabetically_first_word` in `submission.py`.
- [3 points] Implement `euclidean_distance` in `submission.py`.
- [6 points] Implement `mutate_sentences` in `submission.py`.
- [4 points] Implement `sparse_vector_dot_product` in `submission.py`.
- [4 points] Implement `increment_sparse_vector` in `submission.py`.
- [4 points] Implement `find_nonsingleton_words` in `submission.py`.

### Problem 2: 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 [NeurIPSEthical Guidelines](https://neurips.cc/public/EthicsGuidelines) (<https://neurips.cc/public/EthicsGuidelines> ) 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 expected:** 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. [2 points] 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 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. [2 points] 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. [2 points] 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. [2 points] 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.

**BEST!!!**