Assignment Rubric

|  |
| --- |
| **Complete the following for “In Progress”** |
| * Code included in Vectors arithmetic section runs with no error and all outputs are visible * Loading GloVe embeddings with no errors * Similarities and Analogies: run all cells with no errors * Remaining code may run with some errors; answers are attempted but not entirely correct or incomplete |
| **Complete the following for “Good”** |
| * Question 1: 2 out of 3 arithmetic questions are answered correctly (showing work is not necessary) * Question 2 is answered correctly * analogy() is completed and tested on the two examples given * Quantifying Bias in Word Embeddings: run all cells with no errors * Question 3: a list of nurses and doctors is provided, with their relative gender split * Question 5: get() is used to see value of ‘student’ * Question 7: most words are biased toward female (average is not necessary) * Question 9: the student recognizes that separating clusters should be easy; arguments can be presented in support or against the fact that the 2 clusters correspond to the gendered words. * Question 10: code runs and produces correct output (1, or 100%). The student should comment on this result, saying that it shows that clustering is perfectly able to split the words in two groups corresponding to the original gender biases * All cells in “Neutralize and Equalize” run without error * Question 11: the student should recognize that clustering appears to be more difficult (they may also comment that it is possible that it works better at higher dimensions) * Question 12.1: code runs and produces correct output (~72%); the student should recognize that this is an improvement, but a lot of words can still be associated to the original gendered labels. * Question 13: student should identify some differences, especially the improvement in ballerina, which now only includes dance-related words * Question 15: student performs all tasks; acceptable errors include failing to report all required elements (e.g. prompt, name of model), and a shallow evaluation of the responses |
| **Complete the following for “Excellent”** |
| * Question 1: all arithmetic questions are answered correctly (showing work is not necessary) * Question 3: the student comments on how the different gender split of doctors and nurses in Wikipedia entries is resulting in the embedding learning the association between the gender and the profession * Question 4: the student lists the limitations of undersampling and oversampling, and possibly comments on the existence of other biases in the corpus * Question 5: ‘student’ has a negative value, so it is identified as having bias toward woman * Question 6: the student makes a reasonable attempt at evaluating words that should or should not be gendered (responses may vary) * Question 8: the student makes a reasonable attempt at evaluating the similar words to the 3 given (responses may vary) * Question 12.2: the student should explain that looking at the 2D plot was not an effective way to estimate bias, because no clusters were visible, but we found that words could still be grouped by gender with a good accuracy * Question 14:   1. Word embeddings learn their bias from the corpus they are trained on and from the way some words tend to be associated; training on a less biased corpus would have a positive effect and produce a less biased embedding.   2. A lot of possible answers are acceptable, including: translations, search engines, sentiment analysis (with some words having unfair negative or positive connotation), various recommendation systems, recruitment/screening tools, etc.   3. Possible safeguards include disclosing reports on model performance and presence of biases, specifying the recommended context of use, frequent auditing and allowing users to report biased results. * Question 15: student performs all tasks and reports all required elements, including a thorough evaluation of the responses |