**Assignment 4.1 Short Answer Template**

These questions are based on the Assignment 4.1 Exercises document.

**Written Questions**

(Short answer, 2-3 sentences each):

* Using the confusion matrix and report in the Categorical Data for Reviews section, summarize the results you obtained using the balanced and unbalanced perceptron. Report it in terms of both f1 score and accuracy.
* Of all the methods used on the clothing ratings dataset, which of them performs best as a classifier? Would you recommend this system to predict product ratings?

(Short answer, 1-2 sentences each):

* In the clothing ratings, you plotted height and waist size (plots appear in the Assignment 4.1 Exercises document). What transformation could be applied to the waist data, to make it look more like a normal distribution?
* You obtained a plot of accuracy vs. the number of words chosen for count vectorizer and tf-idf vectorizer (or refer to the plot in the last section of the Assignment 4.1 Exercises document). How many features are used to obtain the best accuracy on this dataset?
* In the text classification section, you printed a list of the top ten words. What are they, and do you think these words make sense in the context of clothing reviews?