

# CSE 255, Winter 2015: Homework 3

## Instructions

Please submit your solution **at the beginning of the next lecture (January 26)** or outside of CSE 4102 beforehand. Please complete homework **individually**.

Download the “50,000 book reviews” data, as well as the “5,000 book images” data from the course webpage:  
[http://jmcauley.ucsd.edu/cse255/data/amazon/book\\_descriptions\\_50000.json](http://jmcauley.ucsd.edu/cse255/data/amazon/book_descriptions_50000.json)  
[http://jmcauley.ucsd.edu/cse255/data/amazon/book\\_images\\_5000.json](http://jmcauley.ucsd.edu/cse255/data/amazon/book_images_5000.json)

Code is provided on the course webpage showing how to load and perform simple processing on the data. Executing the code requires a working install of Python 2.7 with the `scipy` packages installed. This week’s code additionally requires that the `sklearn` package be installed.

## Tasks

1. Download the code from the previous lecture from <http://jmcauley.ucsd.edu/cse255/code/lecture2.py>. Modify the SVM code so that it predicts whether a book belongs to the “Literature & Fiction” category from its image. Note that the classifier has the regularization parameter set to  $C = 1000$ . What fraction of examples are correctly labeled (the “classification accuracy”) on the *train* and on the *test* set?
2. (2 marks) Using the same classifier as above, compute the following values (for the *train* and for the *test* set):
  - (a) The number of true positives, true negatives, false positives, and false negatives (1 mark).
  - (b) The true-positive rate and the true-negative rate (1 mark).
3. (2 marks) Use the function `clf.decision_function(X_test)` to rank points in the test set according to the decision function (sim. for the training set).
  - (a) What are the **labels of** the 10 points with the highest scores (1 mark)?
  - (b) What is the `precision@10` (1 mark)?