Title (for quality control)
CS5785 Fall 2019 Final Project

Paper summary
This solution incorporates Partial Least Squares Linear Regression along with ensemble/boosting to predict images based on descriptions. The preprocessed data (i.e. the feature matrix X) included bags of words for description, bags of words for tags, and Sentence2Vec, while the label matrix Y includes the ResNet image feature vectors. The regression model is then trained via cross validation on 8,000 rows, and tested on the other 2,000 rows. The trained regression model then predicts label vectors and compare it with the top 20 label vectors ranked by cosine distance. To optimize for space and efficiency, the team also leverages PCA.

Technical Deep Dive
This paper uses the intermediate ResNet features, and the intuition for that is the intermediate features may provide more useful information regarding the images, as they have not gone through the final neural net layer, which tends to discard specific information for the sake of generalizing. Another good thing they did was using cross-validation, to split their training data into 8000 and 2000. This allows you to fine tune parameters and minimize overfitting.

What Problems were they trying to solve?
This paper attempts to solve the problem of cross-model matching: given different types of input data (text, video, audio, image), use pattern matching to identify the targeted output. To solve relevance measuring problems, they focus their attention toward using the appropriate features to train their regression models.

Critique: improvements
One way this work could be improved is to consider not doing lemmatization, in order to allow distinctions between similar words, such as "skateboarder" and "skateboard." This helped improve our results when we were experimenting.

Originality/Creativity: 4
Writing Quality: 5
Validity/Correctness: 5