Mini Project: Using Keras to analyze IMDB Movie Data

Now, you're ready to shine! In this project, we will analyze a dataset from IMDB and use it to predict the sentiment analysis of a review.

Workspace

To open this notebook, you have two options:

- Go to the next page in the classroom (recommended)
- Clone the repo from **Github** and open the notebook **IMDB_in_Keras.ipynb** in the **imdb_keras** folder. You can either download the repository with `git clone https://github.com/udacity/deep-learning.git`, or download it as an archive file from **this link**.

Instructions

In this lab, we will preprocess the data for you, and you'll be in charge of building and training the model in Keras.

The dataset

This lab uses a dataset of 25,000 **IMDB** reviews. Each review comes with a label. A label of 0 is given to a negative review, and a label of 1 is given to a positive review. The goal of this lab is to create a model that will predict the sentiment of a review, based on the words in the review. You can see more information about this dataset in the **Keras** website.

Now, the input already comes preprocessed for us for convenience. Each review is encoded as a sequence of indexes, corresponding to the words in the review. The words are ordered by frequency, so the integer 1 corresponds to the most frequent word ("the"), the integer 2 to the second most frequent word, etc. By convention, the integer 0 corresponds to unknown words.

Then, the sentence is turned into a vector by simply concatenating these integers. For instance, if the sentence is "To be or not to be." and the indices of the words are as follows: