Logistic Regression with L2 regularization

The goal of this assignment is to implement your own logistic regression classifier with L2 regularization. You will do the following:

- Extract features from Amazon product reviews.
- Convert an dataframe into a NumPy array.
- Write a function to compute the derivative of log likelihood function with an L2 penalty with respect to a single coefficient.
- Implement gradient ascent with an L2 penalty.
- Empirically explore how the L2 penalty can ameliorate overfitting.

If you are doing the assignment with IPython Notebook

An IPython Notebook has been provided below to you for this assignment. This notebook contains the instructions, quiz questions and partially-completed code for you to use as well as some cells to test your code.

Make sure that you are using GraphLab Create 1.8.3. See this post for installing the correct version of GraphLab Create.

What you need to download

If you are using GraphLab Create:

- Download the Amazon product review dataset (subset) in SFrame format. Notice the subset suffix: amazon_baby_subset.gl.zip
- Download the companion IPython notebook: module-4-linear-classifier-regularization-assignment-blank.ipynb
- Download the list of 193 significant words: important_words.json
- If you are using Amazon EC2, download the binary files for NumPy arrays: module-4-assignment-numpy-arrays.npz. See the companion notebook for the instructions.
- Save both of these files in the same directory (where you are calling IPython notebook from) and unzip the data file.

If you are not using GraphLab Create:

- If you are using SFrame, download the Amazon product review dataset (subset) in SFrame format: amazon_baby_subset.gl.zip
- If you are using a different package, download the Amazon product review dataset (subset) in CSV format: amazon_baby_subset.csv.zip
- Download the list of 193 significant words: important_words.json

If you are using GraphLab Create and the companion IPython Notebook

Open the companion IPython notebook and follow the instructions in the notebook.

If you are using other tools

This section is designed for people using tools other than GraphLab Create. You will not need any machine learning packages since we will be implementing logistic regression from scratch. We highly suggest you use SFrame since it is open source. In this part of the assignment, we describe general instructions, however we will tailor the instructions for SFrame.