AMAZON REVIEWS FOR SENTIMENT ANALYSIS

The aim of this project is to build a sentiment analysis model which will allow us to categorize words based on their sentiments, that is whether they are positive or negative.

What is Sentiment Analysis?

Sentiment Analysis is a process of extracting opinions that have different polarities. By polarities, we mean positive, negative or neutral. It is also known as opinion mining and polarity detection. With the help of sentiment analysis, you can find out the nature of opinion that is reflected in documents, websites, social media feed, etc. Sentiment Analysis is a type of classification where the data is classified into different classes. These classes can be binary in nature (positive or negative) or, they can have multiple classes (happy, sad, angry, etc.).

DATASET DESCRIPTION:

This dataset consists of a few million Amazon customer reviews (input text) and star ratings (output labels) for learning how to train fastText for sentiment analysis.

The idea here is a dataset is more than a toy - real business data on a reasonable scale - but can be trained in minutes on a modest laptop.

Content:

The fastText supervised learning tutorial requires data in the following format:

```
__label__<X> __label__<Y> ... <Text>
```

where X and Y are the class names. No quotes, all on one line.

In this case, the classes are `__label__1` and `__label__2`, and there is only one class per row.

__label__1 corresponds to 1- and 2-star reviews, and __label__2 corresponds to 4- and 5-star reviews.

(3-star reviews i.e. reviews with neutral sentiment were not included in the original),

The review titles, followed by ':' and a space, are prepended to the text.

Most of the reviews are in English, but there are a few in other languages, like Spanish.

Source

The data was lifted from Xiang Zhang's Google Drive dir, but it was in .csv format, not suitable for fastText.

Training and Testing

Follow the basic instructions at fastText supervised learning tutorial to set up the directory.

To train:

```
./fasttext supervised -input train.ft.txt -output model_amzn
```

This should take a few minutes.

To test:

```
./fasttext test model_amzn.bin test.ft.txt
```

IMPORTANT: Expect precision and recall of **0.916** if all is in order.

ALL THE BEST