Multi-Label Neural Network on Large Data

1. Project Overview

- **Data**:

  Title & Category - a record in the data is one product sold on Amazon

- **Goal**:

  Predict which categories a product belongs to based on its title

Teamwork:

Data

Titel: Gaok Men's Retro Cotton Multi-Pocket Camo Cargo Shorts

Category: 'Clothing, Shoes & Jewelry' 'Men' 'Clothing' 'Shorts' 'Cargo' 'Cotton'

'Zipper closure' 'Material:100% cotton' 'Zip fly and button closure'

'Multi pockets style' 'Loose and casual style The belt is not included'

'For detailed size information from PRODUCT DESCRIPTION before ordering and choose fits size'

Title: QIBOE Men's Baggy Jeans Denim Sweatpants Loose Pants

Category: 'Clothing, Shoes & Jewelry' 'Men' 'Clothing' 'Jeans' 'Denim'

'Zipper closure' 'Material: cotton' 'Style: hip pop'

'Two side slant pockets and two back pockets' 'Straight fit long pants']

2. Data Processing

1. Iterate through 100,000 records to collect unique words from titles and unique categories:
    - Tokenize titles, then append to a dictionary
    - Count unique categories to a dictionary
    - Filter out low frequency words & categories (N<10)
2. Convert each title & category to a binary vector
    - Using titles and unique token/categories as row and column

	hello	world	how	are	you	alex
hello world	1	1	0	0	0	0
How are you	0	0	1	1	1	0

3. Collect every 20,000 vectors to a matrix (input & output) to train the model

3. Model Training

Multi-Label Classification

- Define a multi-label neural network model:
  - Create model: input (one-hot encoded words), 2 hidden layers (each layer has 200 nodes) and output (one-hot encoded categories): activation function = sigmoid
  - Compile model: loss function = binary_crossentropy; learning rate = 0.05
  - Fit model: batch size = 20,000, epochs = 100
- Training:
  - Train incrementally: every 20,000 records
  - Total training data size = 80,000
  - Training time: 4 minutes
- Testing:
  - Testing data size = 20,000

4. Model Performance

Product title: Fulok Women's Classic Solid Lapel Wool Blend Midi Peacoat Jackets

True Category:

Clothing, Shoes & Jewelry
Women
Clothing
Coats, Jackets & Vests
Wool & Pea Coats
100% Polyester
Imported
Detailed size info,please refer to product description
100% Brand New and High Quality
Made in China
As the screen settings are different,the color may not be the same as the actual
Welcome to buy more products, if you have some question pls contact customer service

Our Prediction: (Top 5)

Clothing, Shoes & Jewelry
Women
Shirt
100% cotton
Imported

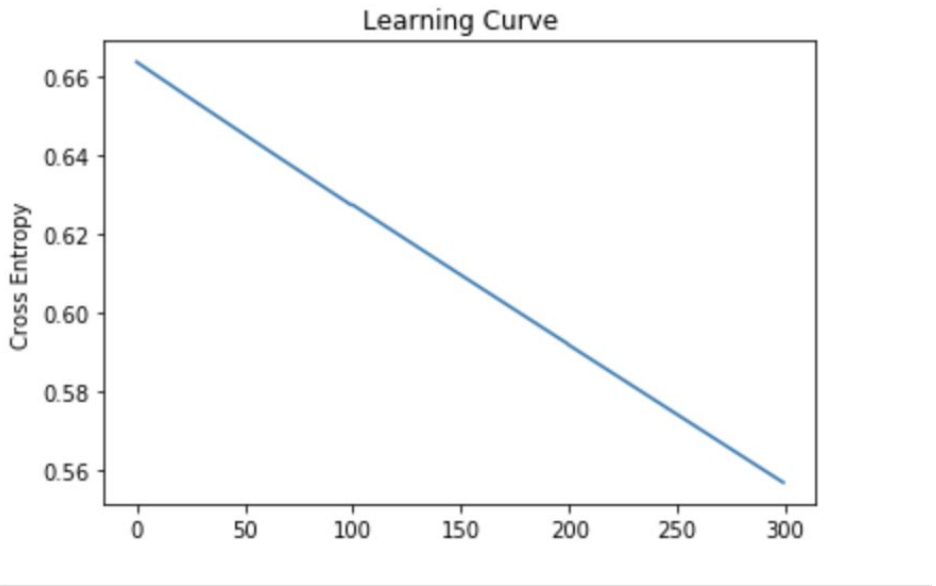
5. Insights

- **Evaluation:**

  Binary Cross Entropy loss: 0.556

- **Learning curve**

  Loss from 0.67 to 0.56



6. Potential Improvements

- Data
  - Customize Category to 20 major categories
    - Multi-Label to Multi--Class
  - Category mapping for each product so the output of model training will be limited to the optimized categories

- Model Tuning
  - Optimize epoch number, batch size, each training size to match each computers' configuration (CPU, RAM, GPU and GPU memory...) to improve training time

QUESTIONS?