EE 622 Advanced Machine Learning
Assignment 3 : Generate text using LSTM

Name: Immidisetti Rakhil
Roll No: 130102026
Department: ECE

Dependencies

Numpy, Scipy, Theano, Keras, Matplotlib.

Running Instructions

- Command for training the model:
  THEANO_FLAGS=device=gpu,floatX=float32,lib.cnmem=0.8,mode=FAST_RUN python final_lstm.py
- The epub version of the file is converted into text file using publicly available epub converters which is used as input for our model.
- For every epoch, the weights are saved as 'lstm_weights.h5' which overwrites file of the previous epoch.

Report

Initial Source: Keras example
https://github.com/fchollet/keras/blob/master/examples/lstm_text_generation.py

Model details:

- First, the unwanted characters like '\n' , '\r' and other redundant hexadecimal characters were removed which reduced the number of characters to predict by half.
- RMSprop optimizer with learning rate as 0.001/0.002 was used.
- The probability scores are transformed by temperature following the below equation.

$$\tilde{p}_i = f_\tau(p)_i = \frac{p_i^{\frac{1}{\tau}}}{\sum_j p_j^{\frac{1}{\tau}}}$$

- For temperature t=1, the function is just the identity function. For $t \to 0$, the freezing function turns sampling into the argmax function, returning the most likely output word. For *t*=0.5, the freezing function is equivalent to squaring the probability of each output word, and then renormalizing the sum of probabilities to 1. The typical perspective is that a temperature like 0.5 is supposed to make the model more robust to errors while maintaining some diversity that you'd miss out on with a greedy argmax sampler.

- Characters are predicted for a given sequence and they are encoded as One-hot bit.
- The text is generated for Temperatures of 0.2,0.5 and 1 and the generated sequence for 0.5 is saved into a text file as it is more robust.

Final model Architecture:

- Two LSTM layers with 128 blocks and the whole ouput is returned for the first layer.
- RMSprop optimizer with learning rate as 0.002, gradient clipping and learning rate decay was used.
- Dropout was 0.2 after every layer.
- Final layer was is a fully connected layer with softmax actiavtion function.
- The sequence is of length 50 and the step for training the model is 5.
- The total parameters of the model was around 2 million corresponding to the 2.5 MB text file as mentioned in the Karpathy's blog.
- Final training loss=**1.2048**

Experiment observations:

- First, started out with one LSTM layer with 128 blocks and observed the loss went down to 1.2 with around 10 epochs
- Then added one more layer with dropout=0.2 after the first layer. The loss decreased to 1.5 after few epochs but then suddenly it spiked to 8.
- I changed the blocks in both the layers and experimented with dropout but obsereved a similar phenomenon.
- Figured out that the gradient is exploding suddenly, so used gradient clipping and learning rate decay. The loss converged to 1.5 around 20 epochs and the decrease after that was sluggish.
- The spike in loss was only observed when the layers or blocks were increased.

Generated text for best model

Temperature = 0.5

1. ----- Generating with seed: "alism which present and the application of the ent"
   alism which present and the application of the entdities offered for sale, cannot appear if the supply of the working of the price to all the position of the price that the same contracted of social actions and the consideration of the laws which being and standard of the market for the absence of the money considerations. the production of the economic calculation and labor is always which the money conducts of capitalism and the nations of socialism which present and the application of the ent
2. ----- Generating with seed: "nstances of a market and as eventy and the and to "
   nstances of a market and as eventy and the and to orces which brought it about are no longer in operation of the best perfect of the conduct of the expansion of the ends under

of the state of the services. the cworch of the entrepreneurs is the process of believe the process of a shorting of the expended by the monopoly prices. the productivity of the prices of all other tool is a very monopoly power to well on the time and the instances of a market and as eventy and the and to

3. ----- Generating with seed: "the banker in a definite monetary part of a orders"
   the banker in a definite monetary part of a ordersexpansion is a progressing system. it produced nor the action of a living of monopoly prices of the production of the exception of the market of the monetary interest of the market proportion with the nature. the expenditure. it is a regulation and works in the employers of the opposition which he cannot be mpistoging in the fact that it is no long as a hegand by the banknote of the production of the banker in a definite monetary part of a orders
4. ----- Generating with seed: "oblems of the market position of the bowrard of sa"
   oblems of the market position of the bowrard of saditions under which we must act today. the cognization of reasoning by the production which a complect of production as the impise that they are not in the sumplee the socialist converting of the better in the production of the consumers for the mode of economic calculation in the appear of the market. the prices of a factors of monopoly in the state of a services of man makes its change in the problems of the market position of the bowrard of sa
5. ----- Generating with seed: "y and which can not helf under of his always becom"
   y and which can not helf under of his always becomoriginary interest nor the amount of further saving of the production of the polical sciences of the products of other is the attainment of exchange of the supply of a fact that the money ready dealt with the rise of more rule that the action of prices depend to production and the govarnment of production and ideal compaiste of the consumers of the outcome of the competition of the rational country and which can not helf under of his always becom