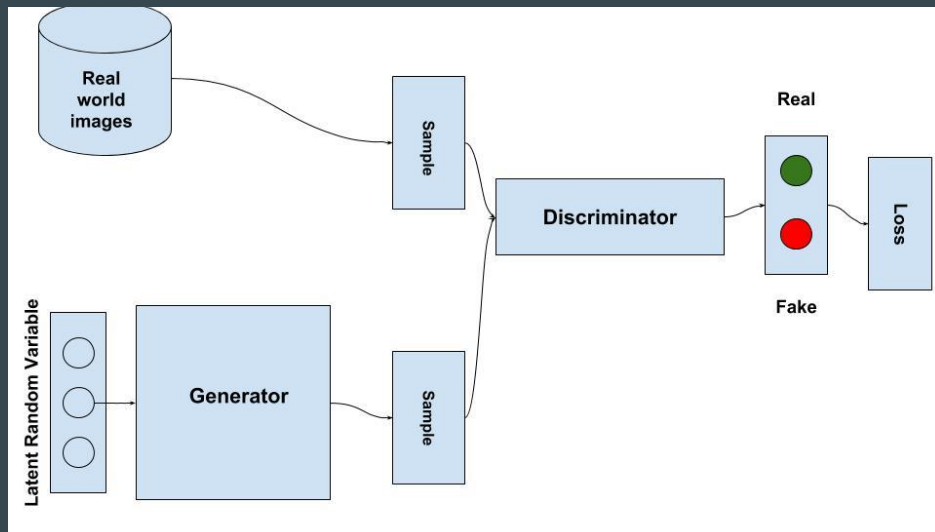# GANs for Word Embeddings

• • •

Akshay Budhkar and Krishnapriya

Introduction

- GANs have shown incredible quality w/ generation of images
- Discrete nature of text makes it harder to train generation of text

GANs for Text

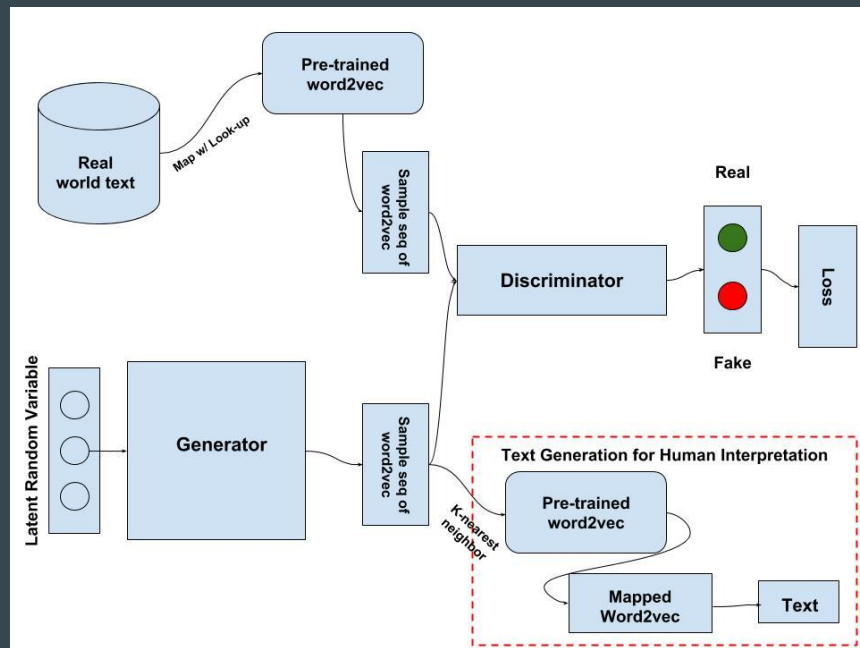Some ways people *approximate* GANs to work for text generation (Goodfellow, 2016)

- Softmax Approximation (Rajeswar, 2017)
- Optimize using Concrete (Kusner, 2016) or REINFORCE (Group in our class)
- Train GANs to generate *continuous* embedding vectors rather than discrete tokens (Ours)

Hypothesis

Training GANs to generate word2vec embedding instead of discrete tokens can produce *better* text because

- Pre-trained real-valued vector space
  - Semantic and syntactic information is embedded in the space itself
- Vocabulary-size agnostic
  - GAN structure can be static when new words are added
  - Variety in text generation due to nature of the embedding space
- No *approximation* needed in the GAN training phase
  - Output of GAN is a word embedding that is fed directly to the discriminator

Figure

$$\min_G \max_D L(D, G) = \mathbb{E}_{x \sim p_r(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

$$= \mathbb{E}_{x \sim p_r(x)}[\log D(x)] + \mathbb{E}_{x \sim p_g(x)}[\log(1 - D(x)]$$

Initial Results

Chinese Poetry Translation Dataset (CMU)

- Replace every first and last word w/ the same characters through the corpus
  - ~100% accuracy after GAN is trained
- Examples of generated sentences
  - <s> i 'm probably rich . </s>
  - <s> can you background anything cream ?
  - <s> where 's the lens . </s>
  - <s> can i eat a pillow ?
  - <s> you can hold the cheeseburger fried </s>
- Learning bi-grams and some tri-grams
- Facing Partial Mode Collapse

<s>will you have two moment ? </s>

<s>i need to understand deposit length . </s>
<s>how is the another headache ? </s>
<s>how there , is the restaurant popular this cheese ? </s>
<s>i 'd like to fax a newspaper . </s>

<s>cruise pay the next in my replacement . </s>
<s>what 's in the friday food ? ? </s>

Future Experiments

- Different architectures & hyperparameter tuning

- Poem-7, Dementia Bank and Newsgroup-20 datasets

- Better metric for quality of text generation
  - Use metrics from the text-translation world

- Performance of conditional variants of our GANs

Thanks!