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
\[
\min_{G} \max_{D} L(D, G) = \mathbb{E}_{x \sim p_{data}} \left[ \log D(x) \right] + \mathbb{E}_{z \sim p_{z}(z)} \left[ \log (1 - D(G(z))) \right]
\]

\[
= \mathbb{E}_{x \sim p_{data}} \left[ \log D(x) \right] + \mathbb{E}_{x \sim p_{data}} \left[ \log (1 - D(x)) \right]
\]
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>
  - <s> where ’s the lens . </s>
  - <s> can i eat a pillow ? </s>
  - <s> you can hold the cheeseburger fried </s>
- Learning bi-grams and some tri-grams
- Facing Partial Mode Collapse

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;s&gt; will you have two moment ? &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; i need to understand deposit length . &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; how is the another headache ? &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; how there , is the restaurant popular this cheese ? &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; i ’d like to fax a newspaper . &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; cruise pay the next in my replacement . &lt;/s&gt;</td>
</tr>
<tr>
<td>&lt;s&gt; what ’s in the friday food ? ? &lt;/s&gt;</td>
</tr>
</tbody>
</table>
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!