

## Questions

### “Training”

1. Describe what it means to use a one-hot embedding representation of a categorical feature.
2. Consider the word “cat”. Let  $v$  be the 50-dimensional GloVe embedding of the word “cat”, and let  $w$  be the 100-dimensional GloVe embedding of the same word. Do the first 50 entries of  $w$  corresponds to the entries of  $v$ ? Why or why not?
3. What is the advantage of `nn.GRU` and `nn.LSTM` over `nn.RNN`?
4. What is the advantage of `nn.GRU` over `nn.LSTM`?
5. Explain why padding is sometimes necessary when training an RNN.

### “Generalization”

1. Consider embedding a tweet by summing the GloVe embeddings of the words that appear in the tweet. Suppose that you would like to measure “how similar” two tweets are, using this tweet embedding. Would you use the cosine similarity or the Euclidean distance?
2. Character-level RNNs are RNNs that takes in a text input one character (one letter) at a time. What are advantages of a character-level RNN (compared to a word-level RNN)? What are disadvantages of a character-level RNN?
3. There are several issues with our implementation of `TweetBatcher`. For one, it is problematic to choose a random minibatch by first randomly choosing a tweet length and then choosing a batch with that sequence length. Why is this?