Account
Dashboard
Courses
Calendar
Inbox
History
Help

Spring 2021

Home
Announcements
Modules
Discussions
Grades
Quizzes
People
UW Libraries
Info and Help
Zoom
UW Resources

Assignment 6: Embeddings, Recurrent Neural Networks, and Sequences (Part 1)

New Attempt

Due May 18 by 5:59pm **Points** 6 **Submitting** a text entry box or a file upload

Suppose we have the following variables ...

```
import numpy as np
Word2Index = { "zero": 0,
               "one": 1,
               "two": 2,
               "three": 3,
               "four": 4,
               "five": 5,
               "six": 6,
               "seven": 7,
               "eight": 8,
               "nine": 9 }
Embedding = np.array([[0.53, 0.80, 0.23],
                      [0.34, 0.48, 0.27],
                      [0.81, 0.19, 0.14],
                      [0.09, 0.43, 0.19],
                      [0.35, 0.75, 0.66],
                      [0.79, 0.37, 0.87],
                      [0.49, 0.53, 0.98],
                      [0.74, 0.30, 0.95],
                      [0.14, 0.31, 0.02],
                      [0.46, 0.52, 0.01]])
sequence = "seven nine two four two"
```

What tensor would be used to represent the input sequence?

Hint: we could encode the sequence as shown below, then compute the product np.matmul(Input, Embedding):

```
Input = np.array([[ 0, 0, 0, 0, 0, 0, 0, 1, 0, 0 ],
                  [ 0, 0, 0, 0, 0, 0, 0, 0, 0, 1 ],
                  [ 0, 0, 1, 0, 0, 0, 0, 0, 0, 0 ],
                  [ 0, 0, 0, 0, 1, 0, 0, 0, 0, 0 ],
                  [ 0, 0, 1, 0, 0, 0, 0, 0, 0, 0 ]])
```

Alternatively, we could use the following:

```
Embedding[[ Word2Index[word] for word in sequence.split(" ") ]]
```

Model 6 (5 points)

Please navigate to the following URL to accept the invitation for this Kaggle task: https://www.kaggle.com/t/285c2913580742ecb9e6bbb578a347b3

Activate the conda environment on your VM:

```
conda activate py37_tensorflow
```

Download the data and create the vocabulary for the newsgroup postings:

```
kaggle competitions download ml530-2021-sp-newsgroups
wget http://nlp.stanford.edu/data/glove.6B.zip
wget https://www.cross-entropy.net/ML530/newsgroups-vocabulary.py.txt
python newsgroups-vocabulary.py.txt
```

Create the tensors for the newsgroup postings:

```
wget https://www.cross-entropy.net/ML530/newsgroups-tensors.py.txt
python newsgroups-tensors.py.txt
```

Run the sample training script:

```
wget https://www.cross-entropy.net/ML530/newsgroups-train.py.txt
python newsgroups-train.py.txt
```

Upload your predictions to kaggle:

```
kaggle competitions submit ml530-2021-sp-newsgroups -f predictions.csv -m "newsgroups submission"
kaggle competitions leaderboard ml530-2021-sp-newsgroups -s
```

◂ Previous Next ▸

Submission

✓ Submitted!
May 18 at 10:46am
Submission Details
Download HW06.txt

Grade: 6 (6 pts possible)
Graded Anonymously: no

Comments:
No Comments

1