Columbia University
IEOR4742 – Deep Learning for OR & FE (Hirsa)
Assignment 6 – Due 12:00 noon on Friday May 8th, 2020

Problem 1 (Differentiable Neural Computers - `DNCs`): In the sample code *train_bAbI.py* for training the `bAbI` dataset we have used the following:

(a) memory matrix size 256×64

(b) number of read heads $R = 4$

(c) input size $X = 159$

(d) clipping gradient by value `tf.clip_by_value` in the range of 10

(e) number of iterations $10,000$

(f) optimizer `tf.train.RMSPropOptimizer` with learning rate of 0.001 and momentum 0.9

Our goal is to assess the performance of the `DNC` architecture under the following cases and compare:

(1) size of memory matrix 256×64, 128×64, 128×32, and 64×32

(2) number of read heads $R = 2, 4, 6$

(3) w/o gradient clipping and w/ gradient clipping by value `tf.clip_by_value` in the range of 5 & 10

(4) optimizer:

- `tf.train.RMSPropOptimizer`
- `tf.train.AdamOptimizer`

(5) number of iterations $20,000$ & $80,000$

There would be 6 scenarios for each group (total of 132 scenarios divided among 22 groups). Each group will be testing `DNC` performance under iterations $20,000$ & $80,000$, w/o and w/ clipping gradients with two different ranges: 6 & 10.

- Groups 1-11 optimizer choice `tf.train.RMSPropOptimizer`
  - Groups 1-3 size of memory matrix 256×64
    - Group 1: number of read heads $R = 2$
    - Group 2: number of read heads $R = 4$
    - Group 3: number of read heads $R = 6$
  - Groups 4-6 size of memory matrix 128×64
    - Group 4: number of read heads $R = 2$
    - Group 5: number of read heads $R = 4$
    - Group 6: number of read heads $R = 6$
  - Groups 7-9 size of memory matrix 128×32
    - Group 7: number of read heads $R = 2$
    - Group 8: number of read heads $R = 4$
    - Group 9: number of read heads $R = 6$
  - Groups 10-11 size of memory matrix 64×32
    - Group 10: number of read heads $R = 2$
    - Group 11: number of read heads $R = 4$

- Groups 12-22 optimizer choice `tf.train.AdamOptimizer`
  - Groups 12-14 size of memory matrix 256×64
    - Group 12: number of read heads $R = 2$
    - Group 13: number of read heads $R = 4$
    - Group 14: number of read heads $R = 6$
  - Groups 15-17 size of memory matrix 128×64
    - Group 15: number of read heads $R = 2$
    - Group 16: number of read heads $R = 4$
    - Group 17: number of read heads $R = 6$
  - Groups 18-20 size of memory matrix 128×32
    - Group 18: number of read heads $R = 2$
    - Group 19: number of read heads $R = 4$
    - Group 20: number of read heads $R = 6$
  - Groups 21-22 size of memory matrix 64×32
    - Group 21: number of read heads $R = 2$
    - Group 22: number of read heads $R = 4$

during testing for each scenario, do on both the original data and the manipulated data i.e. replacing the name of the object by `it` (e.g. replacing `football` by `it` in few places as discussed during the lecture)