IST 597: Foundations of Deep Learning **(Due: 11/13/19)**

Assignment #10000

Instructor: **Dr. Lee Giles** *TAs:* Ankur Mali , K.Zhang *Name: , PSUID:*

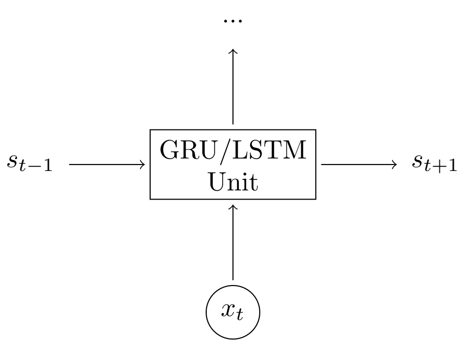
Created by Ankur Mali

Course Policy: Carefully read all the instructions, before you start working on the assignment

- All Problems should be coded in Tensorflow 1.14 or above
- Please typeset your submissions in provided LaTeXtemplate ,give maximum explanation for each subproblems . Please include your name and PSUID with submission along with date and time on the first page.
- Assignments are due at the end of the day at 11:59 pm on the due date given on the webportal Portal.
- No single line answers are accepted in the submission.
- Late assignments will suffer 50 percent loss after the first day and all loss after the second.
- All source materials must be cited. The University Academic Code of Conduct will be strictly enforced.
- We will be creating Canvas submission page for this.You have to submit python file[no ipython allowed] given in gitrepo along with your response pdf.
- All queries related to Assignment should have a subject line IST597:Assignment10000 Queries

Problem 1:Implementing various Recurrent neural network cells using basic tensorflow ops (3+3+2=8 points)

One step of shallow RNN s_{t-1} is previous hidden state, x_t is current input and s_{t+1} is future hidden state. RNNs are stateful models, the process sequence overt time. In the given diagram , RNN takes input at current time step(x_t) and also previous hidden state (s_{t-1}) and then computes output(y_t) and updates the current hidden state(s_t). Our assignment is focused on exploring the effectiveness of various recurrent units in



solving temporal credit assignment problem.

In this assignment we will be implementing and comparing Gated Recurrent Unit with minimal Gated Unit. I have provided the scripts to create LSTM cell in my github repo, please follow it. Current code base can work with keras/Tensorflow eager mode as well as graph mode.

You have to design the forward as well as backward call for your custom RNN cells.[Note:- Feel free to use keras backend with your custom RNN]. People are free to use gradientTape or kerasbackend for backward call.

Below we are providing equations for two Recurrent unit.

1 Gated Recurrent Unit (GRU)

Update equation for Gated Recurrent Unit(GRU). GRU is the simplified version of LSTM and is also as powerful as LSTM in terms of generalization performance. We encourage you to read [1] for in-depth understanding.

$$
\begin{aligned}
\mathrm{z}_t &= \sigma(W_z[s_{t-1}, x_t] + b_z) \\
\mathrm{r}_t &= \sigma(W_r[s_{t-1}, x_t] + b_r) \\
\tilde{\mathbf{s}}_{\mathbf{t}} &= \mathrm{Tanh}(\mathrm{W}_s[r_t \odot s_{t-1}, x_t] + b_s) \\
\mathrm{s}_t &= (1 - z_t) \odot s_{t-1} + z_t \odot \tilde{s}_t
\end{aligned}
$$

2 Minimal Gated Unit (MGU)

Below we have provided the equations for minimal gated unit(MGU). MGU is the simplified version of gated recurrent unit. MGU is 3 times faster than LSTM and 2 times faster than GRU due to less update calls.We encourage you to read [2] for in-depth understanding.

$$
\begin{aligned}
\mathrm{f}_t &= \sigma(W_f[s_{t-1}, x_t] + b_f) \\
\tilde{\mathbf{s}}_{\mathbf{t}} &= \mathrm{Tanh}(\mathrm{W}_s[f_t \odot s_{t-1}, x_t] + b_s) \\
\mathrm{s}_t &= (1 - f_t) \odot s_{t-1} + f_t \odot \tilde{s}_t
\end{aligned}
$$

Things to do

- You will be writing your own training procedure to train RNNCell on notmnist dataset [data]. Report classification error over 3 trials.
- Implement GRU and MGU using custom code provided.
- Compare GRU and MGU over 3 trials. Number of hidden units for the network can vary between 50 to 512. Number of hidden layers can vary between 1 to 4. You have flexibility to choose parameters based on resources available.
- Plot the training vs test curve for all the models. Report your findings.

References

[1] Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).

[2] Zhou, G., Wu, J., Zhang, C., and Zhou, Z. Minimal gated unit for recurrent neural networks. *CoRR abs/1603.09420* (2016).