**Deep Learning Coursework 2**

**WU Zhi 17040772**

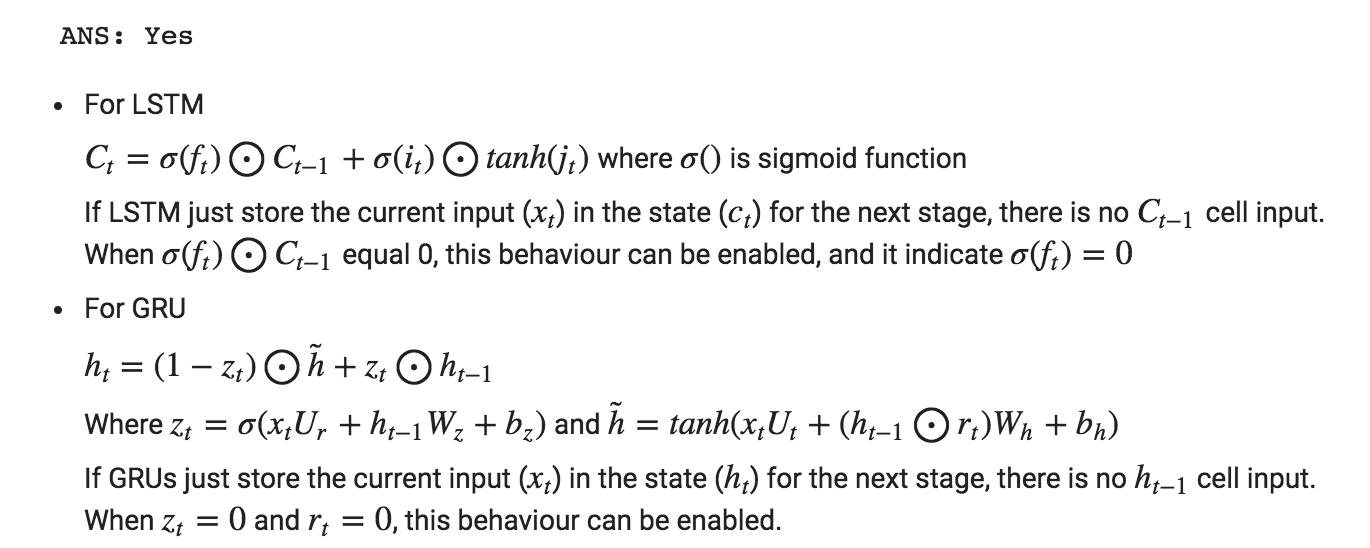
[**UCABWUA@UCL.AC.UK**](mailto:UCABWUA@UCL.AC.UK)

**Department of Computer Science**

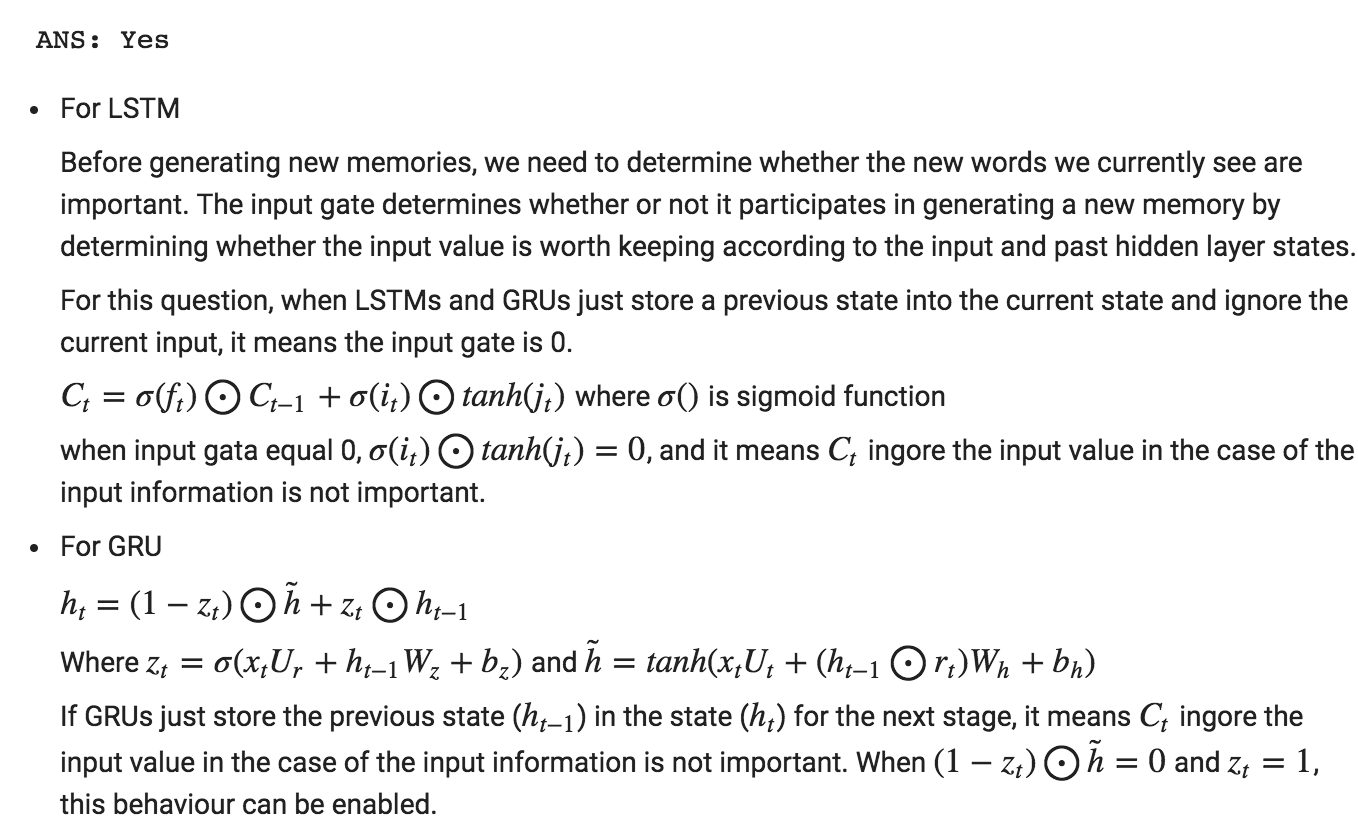
**University College London**

# Q1: Understanding LSTM vs GRU

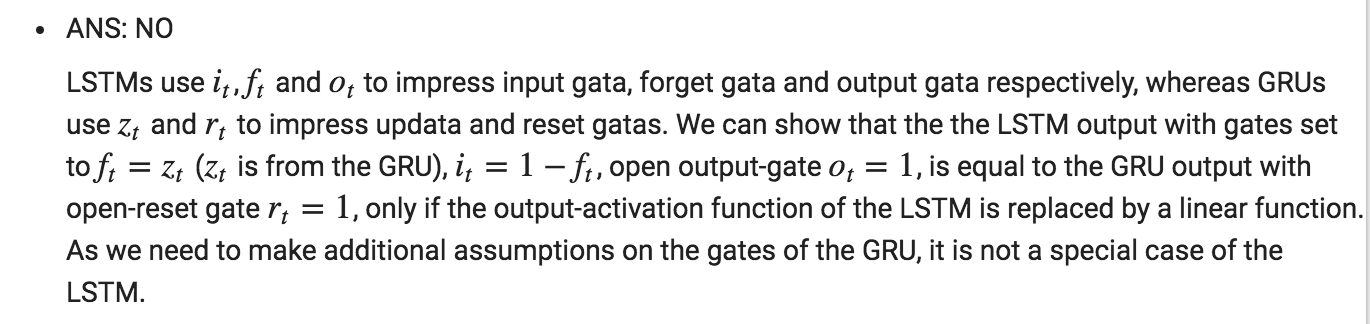
1. Can LSTMs (and respectively GRUs) just store the current input in the state (ct for LSTM and ht for GRU, in the class notation) for the next step? If so, give the gates activation that would enable this behaviour. If not, explain why not. [10 pts]



1. Can LSTMs (and respectively GRUs) just store a previous state into the current state and ignore the current input? If so, give the gates' activation that would enable this. If not, explain why not. [10 pts]

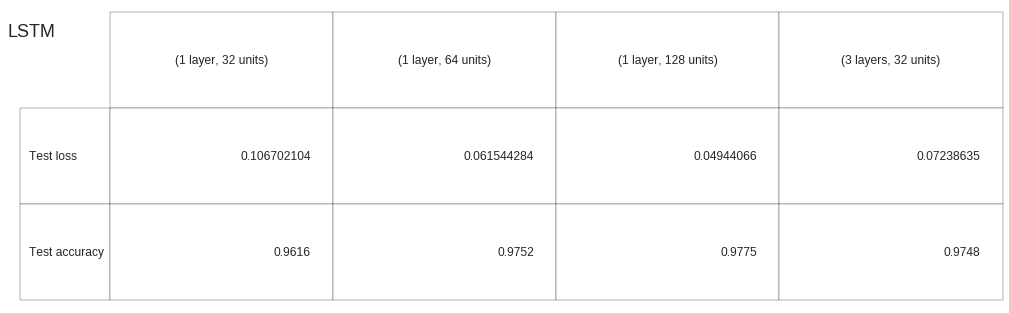


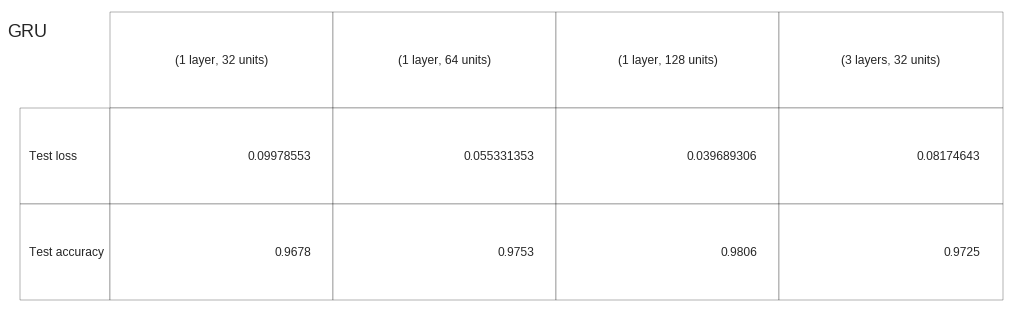
1. Are GRUs a special case of LSTMs? If so, give the expression of the GRU gates in term of LSTM's gates (ot,it,ft). If not, give a counter-example. Assume here the same input. [10 pts]



# Q2: Implementation. Line-by-Line MNIST Classification

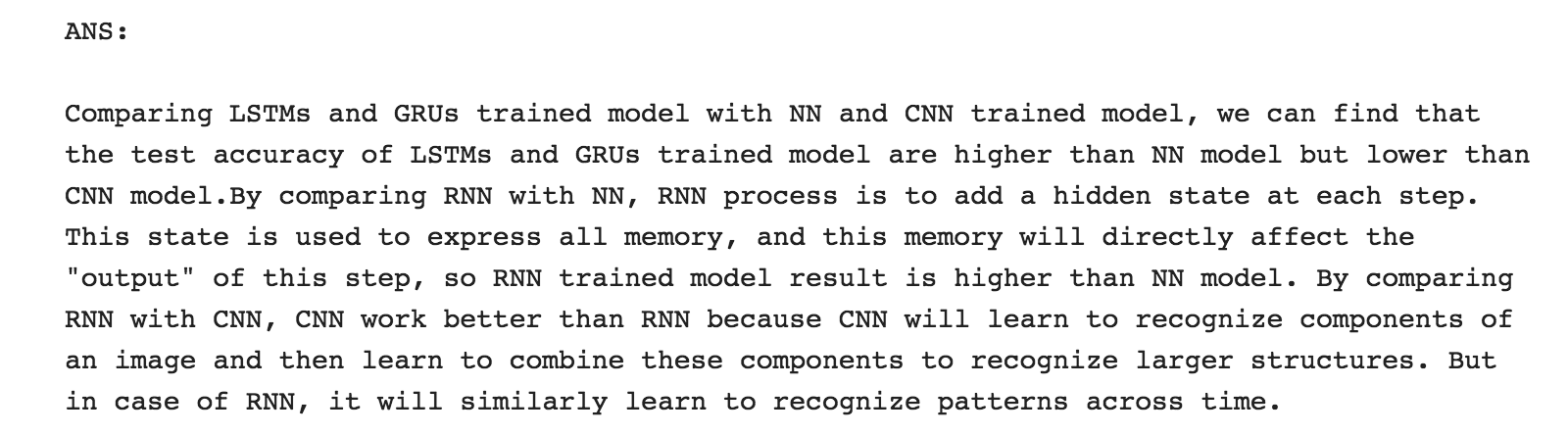
## Result



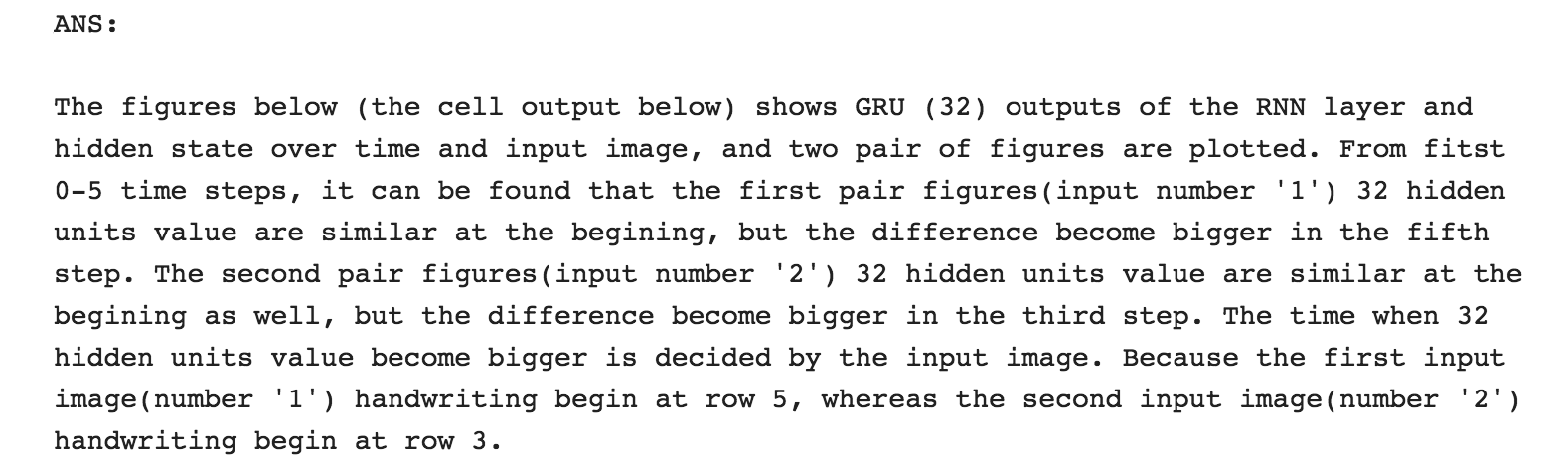


# Q3: Analyse the results

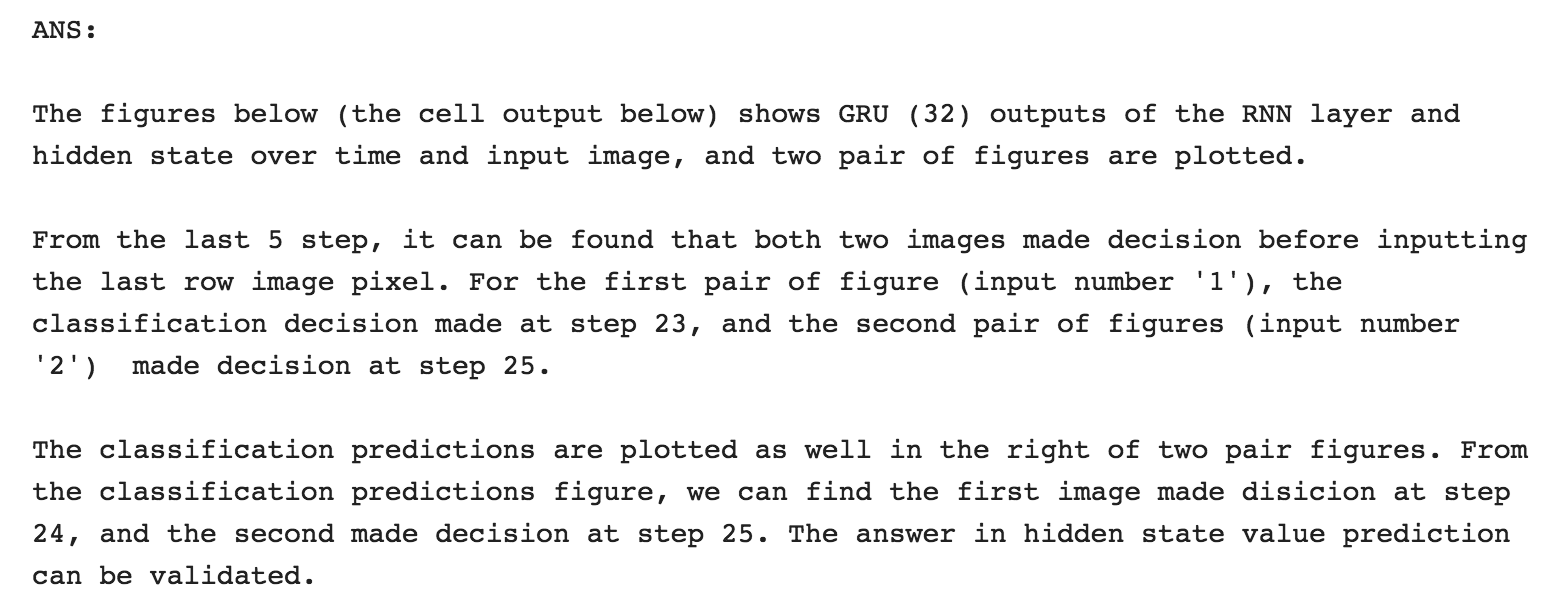
1. How does this compare with the results you obtained in the first assignment(DL1), when training a model that "sees" the entire image at once? Explain differences. [5 pts]



1. Let us take a look closer look at one of the trained models: say GRU (32). Plot the outputs of the RNN layer and hidden state over time. In particular, look at the first 3-5 time steps. Plot the input image along side. You can use python plt.imshow(output\_GRU\_over\_time) for these, where output\_GRU\_over\_time.shape is (T=28,hidden\_units) dimensional. What do you observe? Show at least one pair of these plots to support your observation(s). [5 pts]



1. Now, look at the last 3-5 time steps. What do you observe? When is the classification decision made? To validate your answer to the second part of the question, provide the classification predictions for the last 5 time steps -- that is, pretend whatever the output of the GRU is at that time step is in fact the last output in the computation, and feed that into the classification mapping. [10 pts]



#### **Q3 plot**

