CSC421/CSC2516 Winter 2019 Homework 4

1. LSTM Gradient
2. When we have

So is constant.

1. Multidimensional RNN
2. How many weights does this architecture have? How many arithmetic operations are required to compute the hidden activations?

Num of weights:

Num of operations:

1. How many steps are required to compute the hidden activations?

It requires a total of steps.

For every

It requires one step to compute , one step to add them together and do non-linear activation. To computer we need to computer beforehand, and it requires a total of steps.

1. Give one advantage and one disadvantage of an MDRNN compared to a conv net.

Advantage: MDRNN brings the advantages (taking in time-series information) of RNN to multidimensional data.

Disadvantage: MDRNN requires much more steps to compute so it takes longer to train a MDRNN model.

1. Reversibility

Therefore, the determinant of Jacobian is: