Week 5 – Sequences and RNNs

NOTES FROM FIRST VIDEO

RNNs

In NLP our inputs are word embeddings and our outputs is the probability distribution for the next word or POS label.

The parameters are shared at each time step (the weights and biases)

Unfolding

Again we have labelled, predictions and loss function.

We use loss function separately at each time step and sum up to get total loss.

Backward pass we calculate partial derivatives of Loss function against all parameters.

Backward Propagation Through Time BPTT as we need to back propagate through time (horizontal axis) and the layers (vertical, x, h, y, L\_t and L)

DIAGRAM OF BPTT

U is the ? TODO V is the weight matrix on the inputs (x)

dL\_t/dU = dL/dy\_t x dy\_t/dU so y\_t = f(U h\_t + b\_y)

dL/dW = sum dL\_i/dW so h\_t = f\_h(V x\_t + W h\_t-1 + b\_h)

EQUATION FROM BPTT

As the hidden units from all the previous time staps also depend on V, one needs to go backwards through time to calculate the derivates of the loss w.r.t. V.