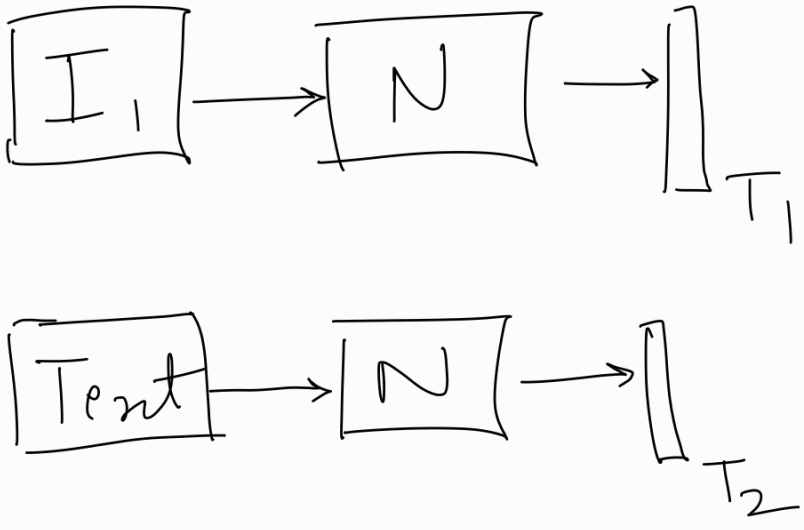
12/3/24

Multi-Modal Self Supervised Learning

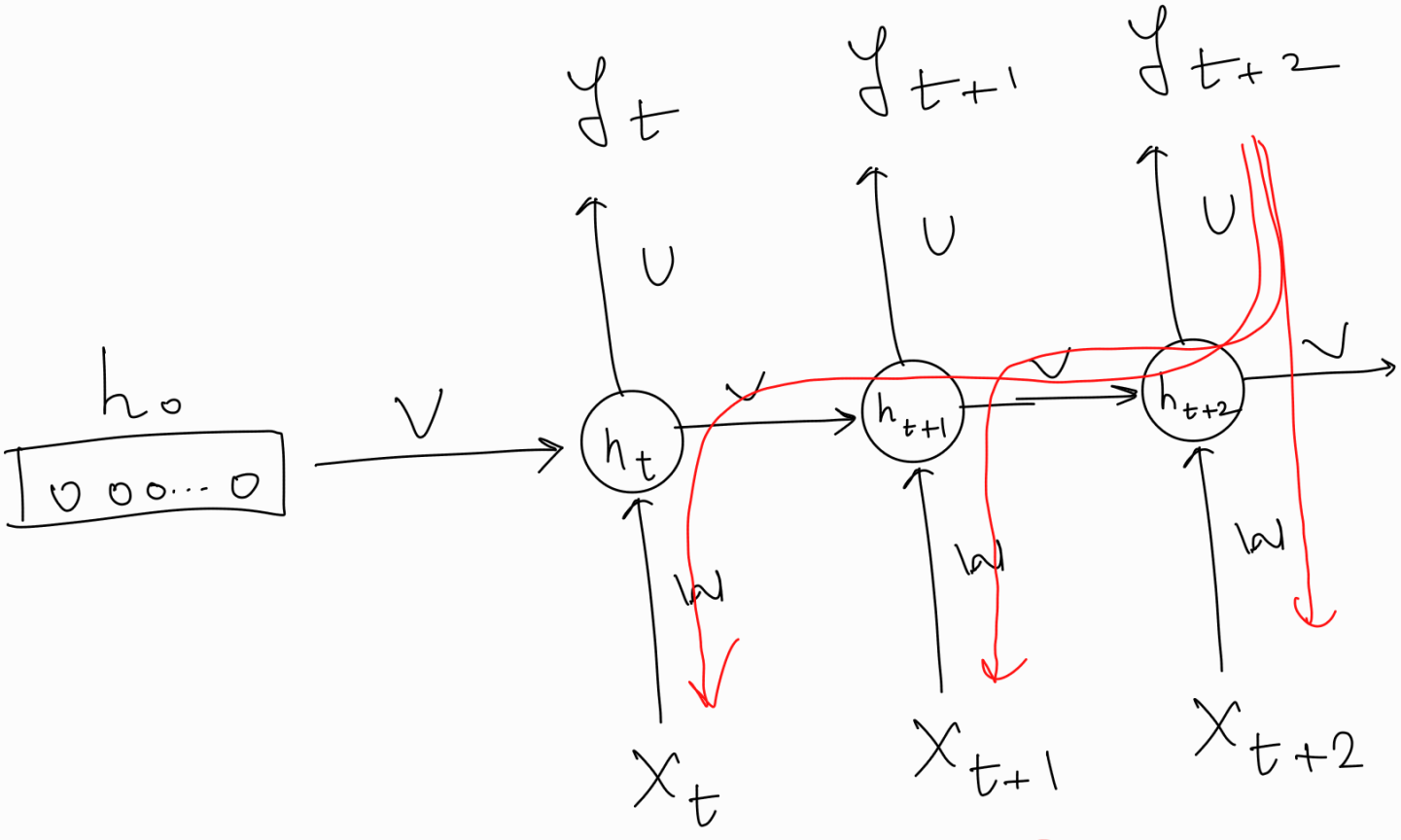


→ Uses: image captioning.

→ Search using keywords to narrow down search space. Now you can apply KNN, etc.

RNNs

Keep content of previous sequence of characters/words.



BPTT: Back Prop Through Time

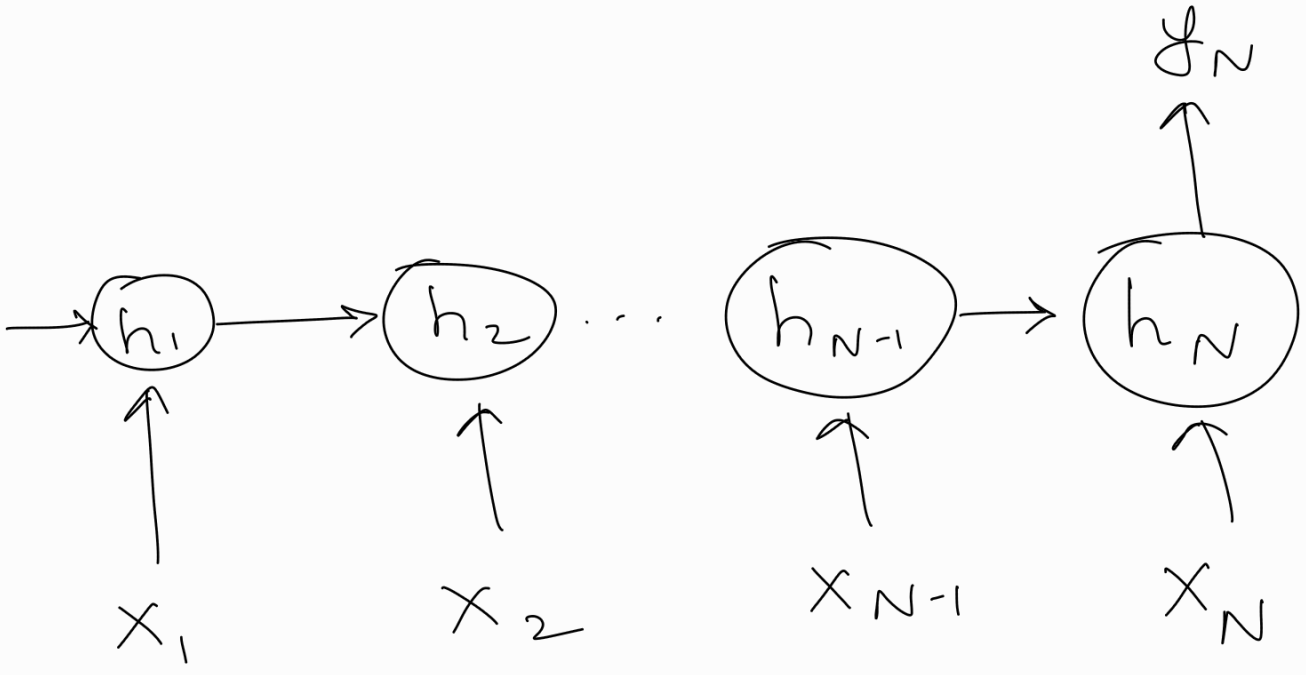
$$h_t = \phi\left(W^T x_t + V^T h_{t-1}\right)$$

$$y_t = \phi\left(U^T h_t\right)$$

Vanishing Gradient problem.

The deeper you go, the lower the effect of o/p on earlier inputs.

Sentiment Analysis Network



Single output

How to embed image?

Pass through CNN, take second-last layer. Use that as 0^{th} content to the RNN (used for image captioning)