Recurrent NNs

NLP - Section 7

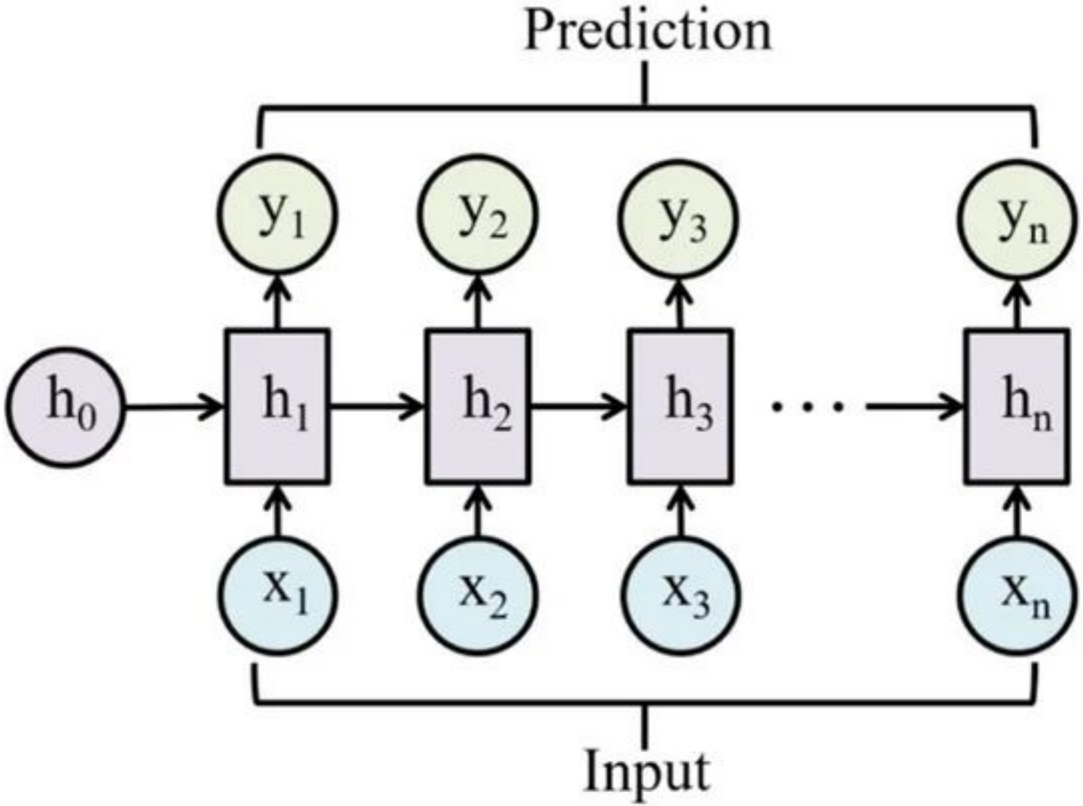
RNNs

- Recurrent means; occurring often or repeatedly.
- The main idea is to overcome the “length” limitation.
- Current LLMs now have the same problem of “length”
- The main idea is to feed the sequential input to the same linear layer over and over.
- We can stack these hidden layers to add complexity to the model.
- Each layer can output y. This y can show us how each input changes the prediction of the model.
- y can also represent prediction of the next x that the model will take. (next-word prediction, sequential weather prediction)

RNNs

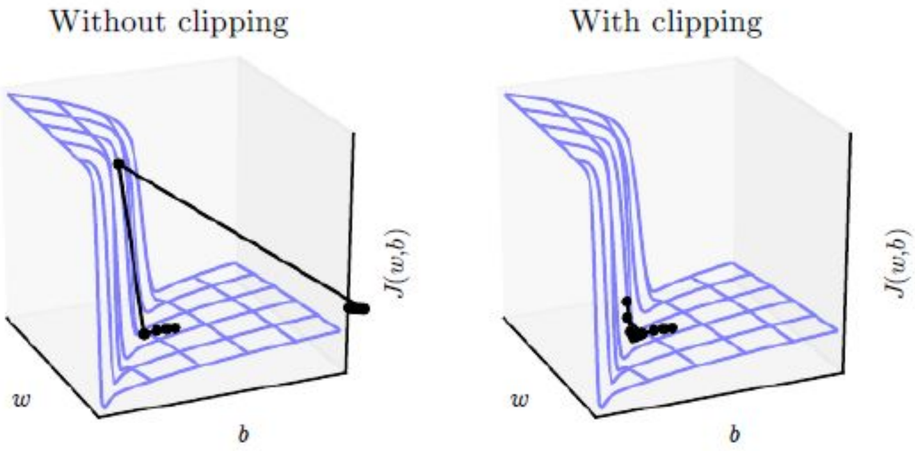
h1, h2, h3 represent the hidden states.

y1 may represent a prediction for x2 or for the true label y.



Exploding Gradients

- If the sequence in the input during training is very large, the W multiplication is raised to the power of n where n is the sequence length
- If n > 1 then we get exploding gradient where the gradient is too large which results in very big changes
- A common solution to this problem is clipping the gradient such that the if the gradient is higher than a certain threshold, it’s normalized.



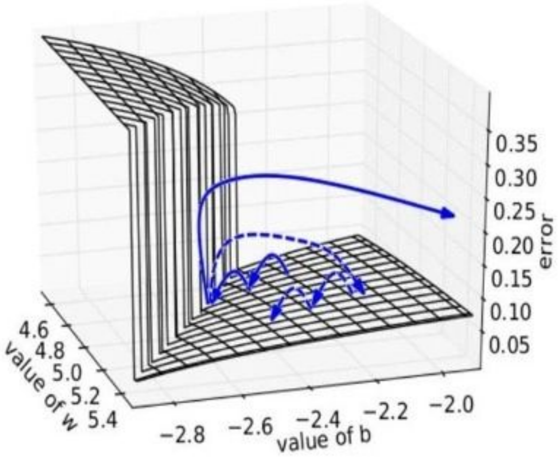
Vanishing Gradients

- If the sequence in the input during training is very large, the W multiplication is raised to the power of n where n is the sequence length
- If n < 1 then we get exploding gradient where the gradient is too small which results in no changes
- Multiple solutions include

  1- Good weight initialization

  2- Batch Normalization

  3- Not using RNN



Code

- RNN demo
- RNN in practice
- Vanishing Gradients
- Exploding Gradients