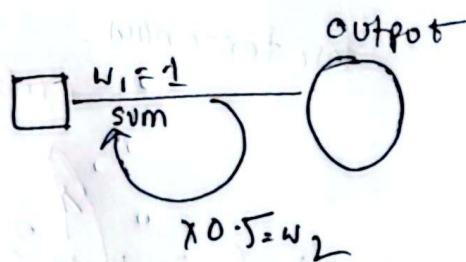


Vanishing and Exploding Gradient Problem

From the equation of Backpropagation, we have known that,

$$\frac{dSSR}{dw_1} = \sum_{i=1}^n -2 (\text{observed}_i - \text{Predicted}) \times [\text{Day before Yesterday} \times w_2^{\sim} + (\text{Yesterday} \times w_2) + \text{Today}]$$

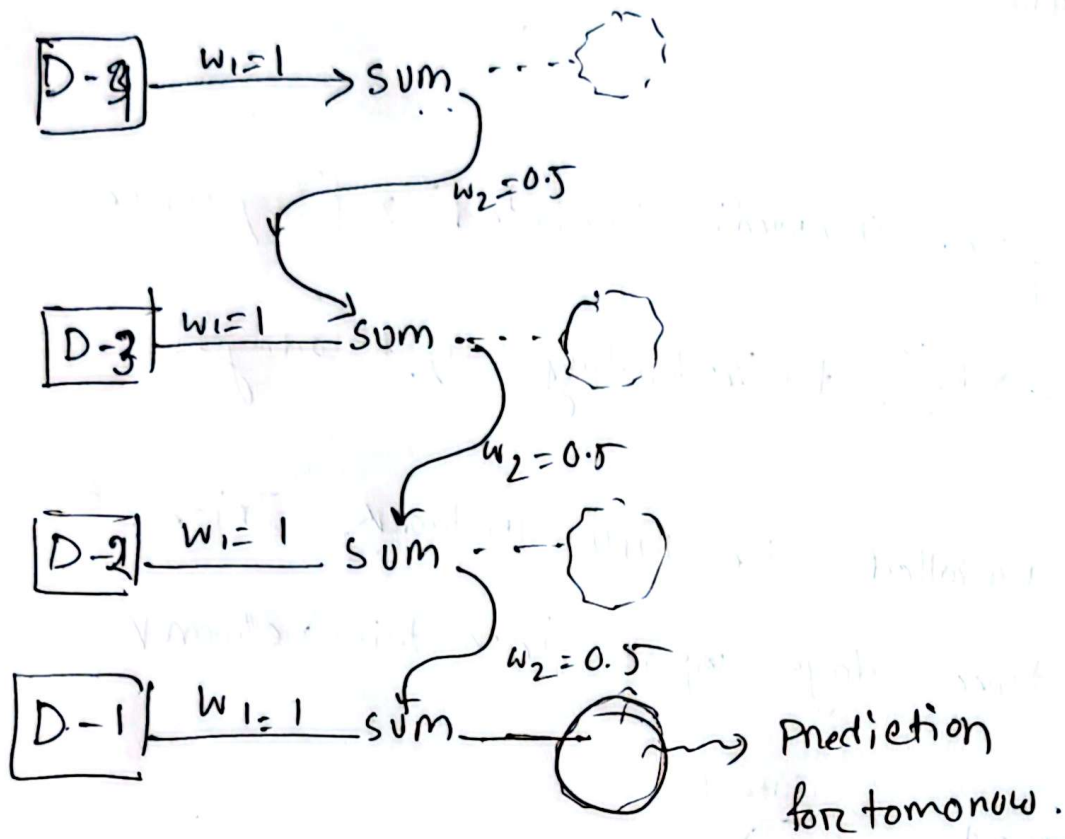
We have unrolled the RNN network ~~Trice~~ ^{pass} to ~~input~~ ^{pass} three days input into this network



That's why the parameter w_2 has square (w_2^{\sim}) in the equation with "Day before yesterday".

However, we will need to deal with many more previous days' data to predict ~~new~~ tomorrow's data.

If we deal with 4 days data, the network will look like



The derivative of SSR \rightarrow

$$\frac{dSSR}{dw_1} = \sum_{i=1}^n -2 (\text{observed}_i - \text{Predicted}_i) \times \left[(D-4 \times w_2^3) + (D-3 \times w_2^2) + (D-2 \times w_1) + D-1 \right]$$

cube of w_2

So, if there are 50 days of data, the input of $D-50$ will be multiplied by

$$w_2^{49}$$

Now, here, $w_2 = 0.5$.

$$\text{So, } \frac{dSSR}{dw_1} = \sum_{i=1}^n -2 (\text{observed}_i - \text{Predicted}_i) \times [(D-50 \times w_2^{50}) + \dots + D_1]$$

$$= \dots \times [(D-50 \times (0.5)^{50}) \dots]$$

↓
This term becomes
very small

Thus the D-50 data gets vanished ~~by~~ ~~multiplied~~
because of getting multiplied by $(0.5)^{50}$.

This is called Vanishing Gradient Problem which
will cause RNN structure to forget older
memory/data. This problem occurs when
 w_2 is between -1 to $+1$.

If $w_2 = 2$, we get -

$$\frac{dSSR}{dw_1} = \sum_{i=1}^n -2 (\text{observed}_i - \text{Predicted}_i) \times [(D-50 \times w_2^{50}) + \dots + D_1]$$

$$= \dots \times D-50 \times (2^{50}) \dots$$

↓
this term becomes
very large.

So, the input of D-50 will get multiplied / scaled by a large number 2^{50} , which will make D-50 data look very significant than D-1's data. So, D-50's data is getting exploded, that's why this problem is called Vanishing Exploding problem.

This problem occurs when W_2 is less than -1 or greater than $+1$.

This problem of vanishing and Exploding Gradient can be solved by Long-Short-term memory ~~net~~ (LSTM) Network

by incorporating long and short term memory feedback along with three types of gates like \rightarrow forget, input and output