RNN (Basic Understanding)

We assume the weekly business health of a firm be determined by 5 financial parameters -

1) Liquidity
2) Profitability
3) Leverage
4) Valuation
5) Solvency.

Let us suppose we are observing such parameters for around 5 weeks. We have

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_5^T \end{pmatrix} = \begin{array}{ccccc} L & P & L & V & S \\ \end{array}$$
$$\begin{pmatrix} 0.5 & 0.3 & 0.6 & 0.7 & 0.2 \\ 0.4 & 0.2 & 0.1 & 0.6 & 0.3 \\ 0.5 & 0.3 & 0.4 & 0.7 & 0.2 \\ 0.6 & 0.5 & 0.8 & 0.7 & 0.4 \\ 0.7 & 0.6 & 0.3 & 0.6 & 0.3 \end{pmatrix}$$

We now consider a matrix of weights $W_{m \times n}$.

$m :=$ Number of hidden units
$n :=$ Number of features in input vector (in our case 5)

Now let's suppose $m = 6$ [we are considering 6 hidden units]

Consequently, W will be of order 6×5
x_1 is of order 5×1

$$U = \begin{pmatrix} x_{11} & u_{12} & u_{13} & u_{14} & u_{15} \\ u_{21} & u_{22} & u_{23} & u_{24} & u_{25} \\ u_{31} & u_{32} & u_{33} & u_{34} & u_{35} \\ u_{41} & u_{42} & u_{43} & u_{44} & u_{45} \\ u_{51} & u_{52} & u_{53} & u_{54} & u_{55} \\ u_{61} & u_{62} & u_{63} & u_{64} & u_{65} \end{pmatrix} \qquad x_1 = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.6 \\ 0.7 \\ 0.2 \end{pmatrix}$$

Let's try to formulate $Ux_1 + b$ where $b = (b_1, b_2, \dots b_6)^T$

$$Ux_1 + b = \begin{pmatrix} u_{11}(0.5) + u_{12}(0.3) + u_{13}(0.6) + u_{14}(0.7) + u_{15}(0.2) + b_1 \\ u_{21}(0.5) + u_{22}(0.3) + u_{23}(0.6) + u_{24}(0.7) + u_{25}(0.2) + b_2 \\ u_{31}(0.5) + u_{32}(0.3) + u_{33}(0.6) + u_{34}(0.7) + u_{35}(0.2) + b_3 \\ u_{41}(0.5) + u_{42}(0.3) + u_{43}(0.6) + u_{44}(0.7) + u_{45}(0.2) + b_4 \\ u_{51}(0.5) + u_{52}(0.3) + u_{53}(0.6) + u_{54}(0.7) + u_{55}(0.2) + b_5 \\ u_{61}(0.5) + u_{62}(0.3) + u_{63}(0.6) + u_{64}(0.7) + u_{65}(0.2) + b_6 \end{pmatrix}$$

We get a vector of shape 6×1.

On simplifying we have $Ux_1 + b = (u_1, u_2, u_3, u_4, u_5, u_6)^T$

We get the state of network for x_1 as

$$\boxed{s_1 = f(Ux_1 + b)}$$

hidden state

Let us suppose, we were using the weekly parameter to predict sales at week t. For 5 weeks, the sales were.

$$\underline{\underline{y}} = \begin{pmatrix} 100 \\ 120 \\ 150 \\ 140 \\ 130 \end{pmatrix}$$

We have $\underbrace{y_1 = \underline{100}}$.

We will need a matrix V of shape $p \times m$.

$p :=$ output features of $\underline{\underline{y_1}}$ $= 1$ $[100, 200 \dots]$

$m :=$ Number of hidden units. $= 6$

Here in our case V will be of shape 1×6.

$V = (V_1, V_2, \dots V_6)$

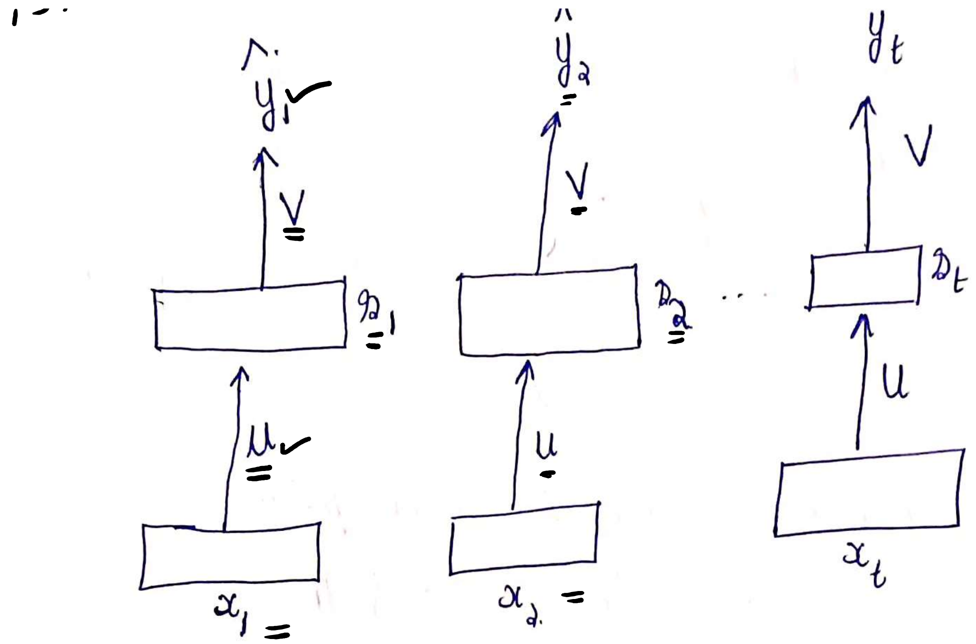
We have $\underbrace{\underline{\underline{V z_1}} + c} = V_1 f(u_1) + V_2 f(u_2) + V_3 f(u_3) + V_4 f(u_4) + V_5 f(u_5) + V_6 \underbrace{f(u_6)} + c.$

The predicted value $= g(V z_1 + c)$

$$\boxed{\underline{\underline{\hat{y}_1}} = g\left[V_1 f(u_1) + \dots + V_6 f(u_6) + c\right]}$$

$$\sum_{i=1}^{K} (y_i - \hat{y}_i)^2$$

$\hat{u}$ $\quad \underline{\underline{\hat{y}_2}}$ $\quad \hat{y}_t$



We formally define a Vanilla RNN as below

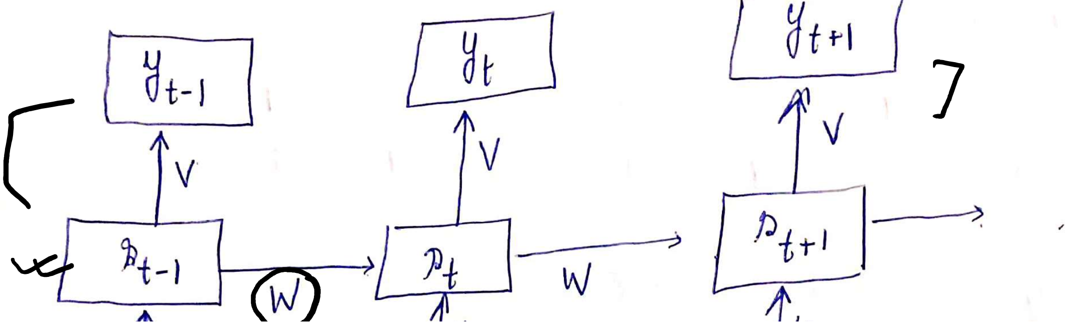
For every x_t we have input to hidden state transition as below -

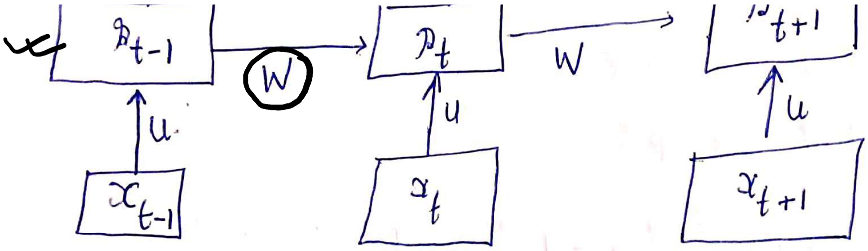
$$s_t = f(Ux_t + b)$$

We have hidden state to output transition as below -

$$y_t = g(Vs_t + c)$$

$\boxed{\text{Parameters to tune :- } U, V, b, c.}$





h_t

Here $s_t = f(Ux_t + Ws_{t-1} + b)$
$y_t = g(Vs_t + c)$ $\rightarrow$ [Parameters to tune: U, V, W, b, c]

$s_t = f(\underbrace{Ws_{t-1}} + Ux_t + b)$

x_t

(3)

A. Consider an RNN where the hidden state equation is $\mathbf{h}_t = \mathbf{W}\,\mathbf{h}_{t-1} + \mathbf{x}_t$. The input vector supplied is $\mathbf{x}_0 = [1, 0]^T$ and the zero vector thereafter, i.e., $\mathbf{x}_1 = \mathbf{x}_2 = \ldots = [0, 0]^T$. The initial hidden state vector is $\mathbf{h}_{-1} = [0\ 0]^T$, and

$$\mathbf{W} = \begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix}$$

Does the hidden state settle down to a particular vector as the number of time steps tends to infinity? If so, what is that vector? Otherwise, explain why it does not settle down to any vector.

We have $h_t = W h_{t-1} + x_t$

Here; $x_0 = \begin{pmatrix}1\\0\end{pmatrix}$ & $x_1 = x_2 = \cdots = \begin{pmatrix}0\\0\end{pmatrix}$

$h_{-1} = \begin{pmatrix}0\\0\end{pmatrix}$ & $W = \begin{pmatrix}0.75 & 0.25\\0.25 & 0.75\end{pmatrix}$

Here; $h_0 = W h_{-1} + x_0$

$= W\begin{pmatrix}0\\0\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\0\end{pmatrix} = x_0 \Rightarrow \boxed{h_0 = x_0}$

$h_1 = W h_0 + x_1$ $\left[\because x_1 = \begin{pmatrix}0\\0\end{pmatrix}\right]$

$= W h_0$

$\boxed{h_1 = W x_0}$

Now, $h_2 = W h_1 + x_2$

$= W(W h_0)$

$\boxed{h_2 = W^2 h_0}$

Now, $h_3 = W h_2 + x_3$

$= W^3 h_0$

$\boxed{h_3 = W^3 x_0}$

In general; $\boxed{h_n = W^n x_0 \quad \text{for } n \geq 1}$

We need to find $\lim_{n \to \infty} h_n$. $\boxed{h_n = W^n x_0}$

Here, $\lim_{n \to \infty} h_n = \lim_{n \to \infty} W^n x_0$

We have $W = PDP^{-1}$ [Diagonalizability Theorem]

where P is a matrix of eigen vectors

D is a matrix with eigen values in diagonal elements

If we have such a representation;

$$\boxed{W^n = PD^n P^{-1}}$$

We will have to calculate eigen values of W first.

The characteristic equation is $|W - \lambda I| = 0$

$$\lambda^2 - (tr(W))\lambda + Det(W) = 0$$

$$\Rightarrow \lambda^2 - (1.5\lambda) + 0.5 = 0$$

Here, $\lambda = \frac{1.5 \pm \sqrt{1.5^2 - 4(0.5)}}{2}$

$$\Rightarrow \boxed{\lambda_1 = 0.5} \quad \boxed{\lambda_2 = 1}$$

$0.5^2 = 0.25$

$0.5^3 = 0.125$

$0.5^n \longrightarrow 0$

$1^{100} = 1$

Here $D = \begin{pmatrix} 0.5 & 0 \\ 0 & 1 \end{pmatrix}$

$$D^n = \begin{pmatrix} 0.5^n & 0 \\ 0 & 1^n \end{pmatrix}$$

$$\boxed{\lim_{n \to \infty} D^n = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}} \quad \left[\because \lim_{n \to \infty} 0.5^n = 0\right]$$

For $\lambda_2 = 1$ we have

$$Mv = \lambda_2 v \quad \text{where } v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$\begin{pmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 0.75 v_1 + 0.25 v_2 - v_1 \\ 0.25 v_1 + 0.75 v_2 - v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -0.25 v_1 + 0.25 v_2 \\ 0.25 v_1 - 0.25 v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -v_1 + v_2 \\ v_1 - v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

We eventually get $v_1 = v_2$. We assume $v_1 = v_2 = 1$

for $\lambda_2 = 1$ we get $\underline{v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}}$ as eigen vector

For $\lambda_1 = 0.5$ we have a $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ such that

$$\begin{pmatrix} 0.75 v_1 + 0.25 v_2 - 0.5 v_1 \\ 0.75 v_2 + 0.25 v_1 - 0.5 v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$0.25 \begin{pmatrix} v_1 + v_2 \\ v_1 + v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

we get $v_1 = -v_2$. We assume $\underline{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Now; $\tilde{P} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$

$\underline{\underline{P}} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$ [Unit Normalization]

Now $W^n = P D^n P^{-1}$

$$= \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0.5^n & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Now $\lim\limits_{n\to\infty} W^n = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$

$$= \begin{pmatrix} 0 & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

$$= \underline{\underline{\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}}} \leftarrow \quad \lim_{n\to\infty} h_n = \underbrace{W^n}_{\lim_{n\to\infty} W^n} \underbrace{x_0}$$

Now $\lim\limits_{n\to\infty} W^n x_0 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}$

$$\Rightarrow \boxed{\lim_{n\to\infty} h_n = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}}$$