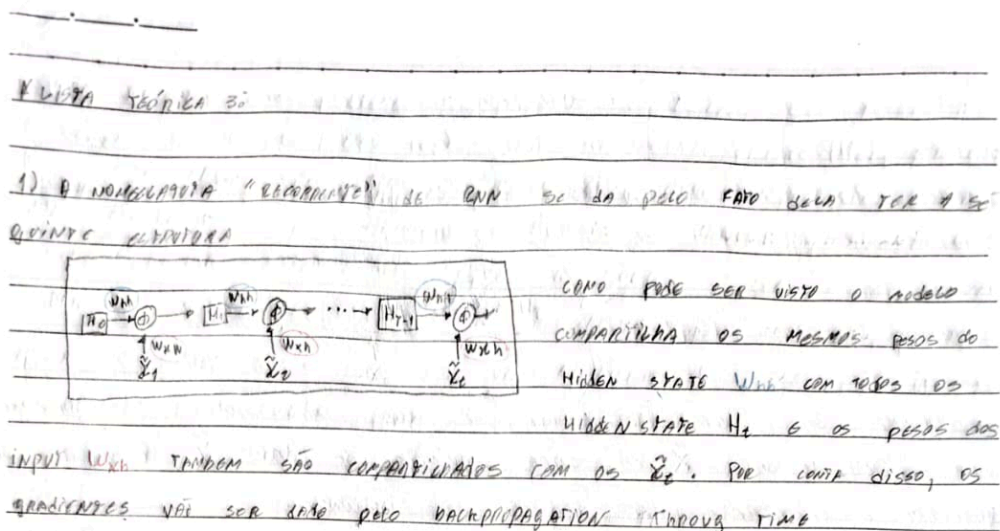


Discente:

- Gabriel D'assumpção de Carvalho - gdc2@cin.ufpe.br

Curso:

- Ciências Atuariais - 7º Período



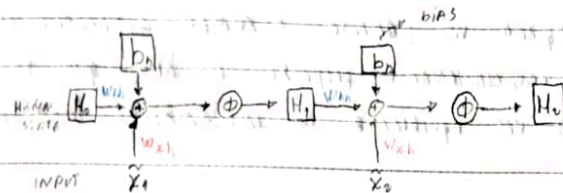
$$\begin{aligned} \frac{\partial L}{\partial w_{11}} &= (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_1}) \cdot (\frac{\partial H_1}{\partial w_{11}}) + (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_2}) \cdot (\frac{\partial H_2}{\partial w_{11}}) + (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_3}) \cdot (\frac{\partial H_3}{\partial w_{11}}) + \dots \\ &= (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_1}) \cdot (\frac{\partial H_1}{\partial w_{11}}) + (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_2}) \cdot (\frac{\partial H_2}{\partial w_{11}}) + (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_3}) \cdot (\frac{\partial H_3}{\partial w_{11}}) + \dots \\ &= \sum_{i=1}^I (\frac{\partial L}{\partial \sigma}) \cdot (\frac{\partial \sigma}{\partial H_i}) \cdot (\frac{\partial H_i}{\partial w_{11}}) \end{aligned}$$

Dividir os aumentos da multiplicação do gradiente a cada etapa de tempo, de modo que seja menor que 1 a cada multiplicação o valor vai ficando próximo de zero, evitando assim que os parâmetros treinados na RNN sofram muitas alterações a cada época, fazendo com que o modelo não chegue em uma convergência, tendo uma loss alta.

PARA MITIGAR A VANISH GRADIENT POSE AUMENTAR A TAXA DE APRENDIZAGEM, USAR LINGUAGEM DE ATIVIDADE RÍGIDA OU VITÁVELS AQUISIÇÕES MAIS MODERNAS COMO LONG SHORT-TERM MEMORY (LSTM) E GATED RECURRENT UNITS (GRU), QUE SÃO AQUISIÇÕES POSSUEM MECANISMO DE MEMÓRIA DE LONGO E CURTO PRAZO, AJUDANDO A LER EM PRESERVAR AS INFORMAÇÕES E EVITAR O VANISH GRADIENT.

2) Abaixo, por ser visto as arquiteturas de uma RNN tradicional, LSTM e GRU.

• RNN TRADICIONAL - MANY-TO-ONE:



\oplus : SOMA

ϕ : FUNÇÃO DE ATIVAÇÃO

\ominus : SUBTRAÇÃO

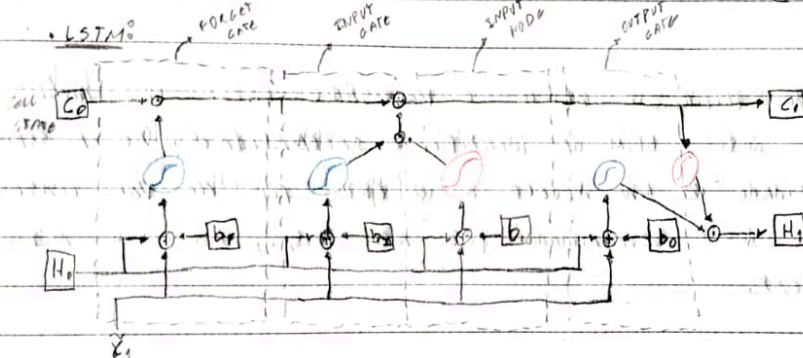
\odot : MULTIPLICAÇÃO

$1-$: COMPLEMENTAR

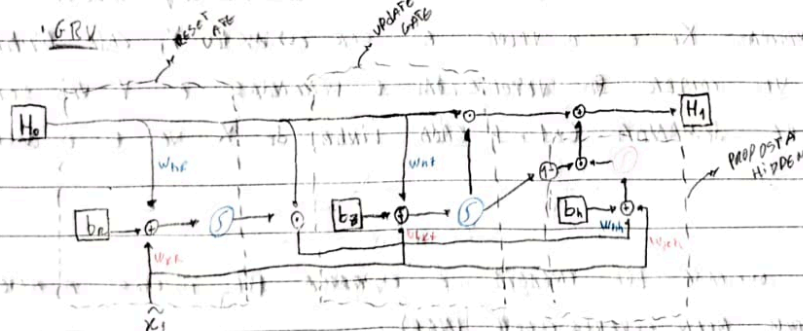
σ : FUNÇÃO SIGMOID

\tanh : FUNÇÃO TANH

• LSTM



• GRU



COMO PODE SER VISTO ACIMA O MODELO DE RNN TRADICIONAL AJUSTA O SEU HIDDEN STATE (H) A CADA PASSO DE TEMPO, FAZENDO COM QUE SOFRE ESQUECIMENTO QUANDO SE VIZITA UMA GRANDE JANELA DE TEMPO.

PARA RESOLVER ESSE PROBLEMA FOI PROPOSTO O LSTM, QUE POSSUI A MEMÓRIA DE LONGO PRAZO (C_t) E UMA DE CURTO PRAZO (H_t), POR ENTO DISSO ESSA ARQUITETURA CONSEGUE RETER INFORMAÇÕES IMPORTANTE DURANTE UMA ESCALA DE TEMPO MAIOR, RESOLVENDO O VANISH GRADIENT. EM CONTRA PARTIDA, ESSE MODELO REQUER UM GRANDE ESFORÇO COMPUTACIONAL. PARA RECEBER A EFICIÊNCIA COMPUTACIONAL FOI PROPOSTO O GRU, QUE TEM APENAS O HIDDEN STATE MAS COM UM EQUILÍBRIO DE INFORMAÇÕES IMPORTANTE DE LONGO E CURTO PRAZO.

3)

① ANTES DE PROPOR UM MODELO É NECESSÁRIO FAZER UMA ANÁLISE EXPLORATÓRIA DOS DADOS, PARA DE IDENTIFICAR SAZONALIDADE. SE A SÉRIE CONTER SAZONALIDADE UM BOM MODELO É O GRU, POIS ELE VAI CONSEGUIR RETER INFORMAÇÕES DESSA SAZONALIDADE POR MAIS TEMPO, ALÉM DE MITIGAR O VANISH GRADIENT.

② OS DADOS PODEM SER ARMazenados EM UM DATAFRAME, CONTENDO UMA COLUNA DAS DEMANDAS X_t E O VALOR A SER ESTIMADO Y , CADA LINHA DA COLUNA X VAI CONTER UM VETOR COM 6 DEMANDAS E Y VAI SER UM ESCALAR DA DEMANDA $t+1$. A CADA LINHA DE X O 6 É DESCRITO DE UM TEMPO.

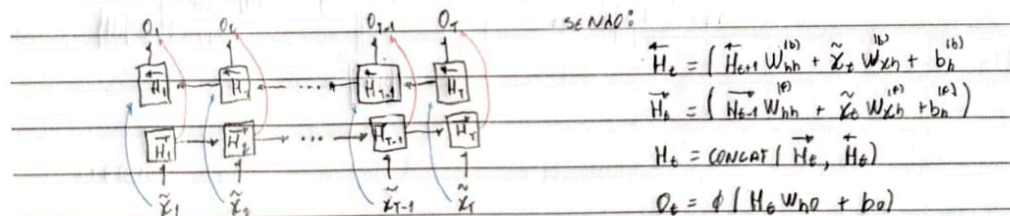
③ COMO A DEMANDA POR ENERGIA É EXPRESSA POR NÚMEROS, A AVALIAÇÃO PODE SER FEITA PELA MÉDIA QUADRA DO ERRO (MSE)

$$MSE = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{y}_t)^2$$

4) Como mencionado na primeira questão, o backpropagation é um somatório dos gradientes de cada passo de tempo, ENTRETANTO A RESPA DA CADA PASSADA UMA MODIFICAÇÃO A CADA TEMPO, PODENDO CAUSAR O EXPLOSION GRADIENT. PARA RESOLVER ISSO FOI PROPOSTO O TRUNCATING BACKPROPAGATION THROUGH TIME, EM QUE ELE LIMITA O GRADIENTE ATÉ O PASSO DE TEMPO PARA UM TAMANHO (L), TERMINANDO A SOMA EM $\partial H_{t-L} / \partial W$.

5) QUANDO TEMOS PROBLEMAS EM QUE O CONTEXTO TEMPORAL É IMPORTANTE, ONDE O PASSO t TAMÉM INFLUENCIA O PASSO $t-1$, COMO EM TRADUÇÃO AUTOMÁTICA E EM MUITOS PROBLEMAS DE NATURAL LANGUAGE PROBLEM (NLP) AS RNN SÃO MUITO UTILIZADAS.

A ESTRUTURA DE UM BI-RNN FOI PROPOSTA PARA QUE UM OUTPUT O_t SEJA INFLUENCIADO PELAS VARIÁVEIS LATENTES H_{t-1} , H_{t+1} E A VARIÁVEL \tilde{x}_t . ABAIXO PODEMOS VER SUA ARQUITETURA



6)

