

Trabalho da Disciplina de Redes Neurais

Prof. Leonardo Mendonza
prof.leonardo@ica.ele.puc-rio.br

Componentes do Grupo:

Nome	Matrícula	e-mail
Flávia Szczerbacki	192.110.229	sflavia@gmail.com
Mariana Gonçalves Almeida Pinto	192.110.208	marianagoncalvesap@gmail.com
Renata Regina da Fonseca Santos	192.110.195	renatarfsantos@zipmail.com.br
Sérgio Gustavo Souza Guerra	201.110.052	sergiog.guerra@gmail.com

Sumário

1	INTRODUÇÃO	3
2	AVALIAÇÃO DA RMSE PARA AS MUDANÇAS SOLICITADAS.....	4
2.1	MÉTRICA RMSE PARA A CONFIGURAÇÃO ORIGINAL	4
2.2	MUDAR O OTIMIZADOR ADAM PARA SGD DENTRO DA CONFIGURAÇÃO DA REDE	4
2.3	VOLTAR A CONFIGURAÇÃO INICIAL E MUDAR O NÚMERO DE NEURÓNIOS DA CAMADA LSTM	5
2.4	COLOCAR UMA SEGUNDA CAMADA LSTM:.....	6
2.5	MUDAR O NÚMERO DE ÉPOCAS.....	6
3	COM AS MUDANÇAS FEITAS PROPONHA SUA CONFIGURAÇÃO IDEAL	8
4	COMPARAÇÃO DOS RESULTADOS UTILIZANDO GRU	9
5	RESULTADOS ALCANÇADOS EM CADA UMA DAS MUDANÇAS.....	10
6	CONCLUSÃO	12

1 Introdução

O objetivo do trabalho é fazer previsões para séries temporais com redes LSTM, utilizando uma base de pessoas mortas por doenças pulmonares na Inglaterra. Cada previsão será feita seguindo as mudanças solicitadas para o notebook `RedesRecorrentespul.ipynb` (previamente fornecido), utilizando como principal métrica de avaliação a RMSE (Root Mean Squared Error).

A métrica RMSE é uma medida de distância entre o destino numérico previsto e a resposta numérica real. Quanto menor o valor de RMSE, melhor será a precisão preditiva de um modelo. Retorna a magnitude do erro e não sua direção, um modelo com previsões perfeitamente corretas, teria uma RMSE igual a 0.

A RMSE tem a propriedade de estar na mesma escala que a variável de resposta e compreender o contexto da aplicação no mundo real dos dados com os quais se está trabalhando, é essencial para interpretá-la. É uma boa medida de quão precisamente o modelo prevê a resposta e é o critério mais importante para ajuste, se o objetivo principal do modelo for a previsão.

A MAPE (Mean Absolute Percent Error) será utilizada como segunda métrica para avaliação do modelo. Expressa a precisão como uma porcentagem do erro, por isso seu entendimento pode ser mais fácil do que as outras estatísticas de medida de precisão. Por exemplo, se a MAPE for 5, em média, a previsão está errada em 5%.

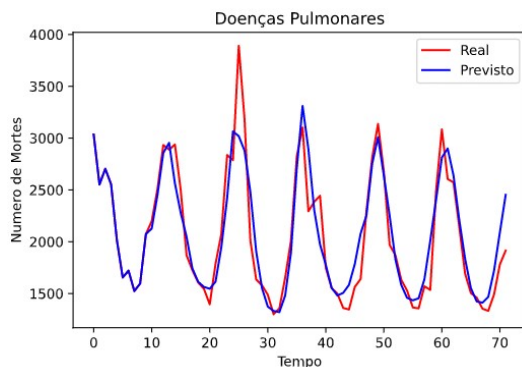
2 Avaliação da RMSE para as Mudanças Solicitadas

Nesta seção faremos a avaliação da RMSE para cada uma das mudanças solicitadas.

2.1 Métrica RMSE para a configuração original

1. Configuração Original

RMSE	MSE	MAPE	MAE
247,41615	61214,7503	10,19675%	195,67470



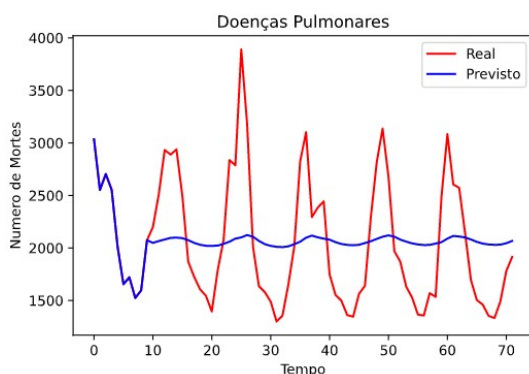
O modelo original apresenta a métrica RMSE de, aproximadamente, 247. Isto significa que o modelo erra a previsão em 247 mortes.

Apresenta a métrica MAPE com valor de aproximadamente, 10,2%, isto é, a previsão está errada em 10,2%.

2.2 Mudar o otimizador *adam* para *sgd* dentro da configuração da rede

2. Mudar o otimizador de adam para sgd dentro da configuração da rede:

RMSE	MSE	MAPE	MAE
538,44489	289922,9023	27,37066%	478,54623



Ao mudar o otimizador *adam* para *sgd*, a precisão do modelo diminuiu, com a RMSE de 538 e MAPE de 27,4%, aproximadamente.

O otimizador SGD, em sua configuração padrão, não apresenta bom desempenho para treino de redes recorrentes como a LSTM. Não utiliza momentum e, especificamente para esse modelo, a taxa de aprendizado padrão 0,01 mostrou-se pequena para o treinamento com 250 épocas.

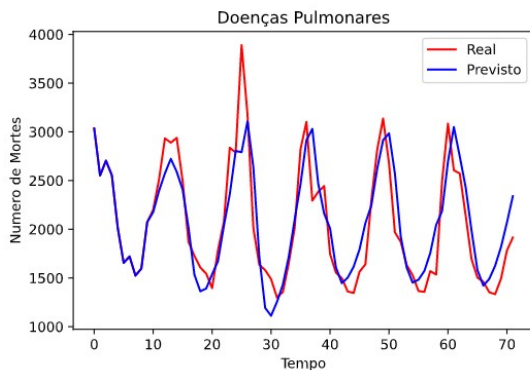
O artigo [Sutskever et al., 2013](#), demonstra que, embora o otimizador SGD não apresente bom desempenho para treino de redes recorrentes como a LSTM, esta lacuna pode ser eliminada ou quase eliminada pelo uso cuidadoso de métodos clássicos de momentum ou gradiente acelerado de Nesterov (NAG). A utilização do otimizador SGD, associado a um valor alto para o Nesterov momentum, teve um

efeito dramático no desempenho do modelo, embora o resultado não tenha se igualado à do otimizador Adam.

Diferente do otimizador SGD que, em sua configuração padrão, não descrece a taxa de aprendizado durante o treinamento, o adam é um método de descida de gradiente estocástico que suporta taxas de aprendizagem adaptativas, método que acelera o treinamento e simplifica a configuração da rede neural.

3. Mudar o otimizador de adam para sgd, utilizando Nesterov momentum e learning rate 0,1:

RMSE	MSE	MAPE	MAE
293,13725	85929,44749	13,93915%	264,62549



Para melhorar o modelo utilizando o otimizador SGD, após alguns experimentos chegamos aos hiper parâmetros abaixo:

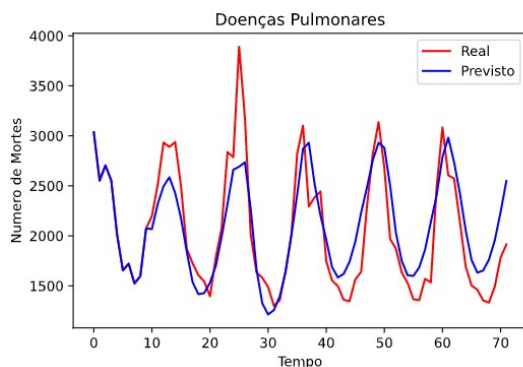
Learning Rate = 0,1
Decay = Learning Rate/n° de épocas = 0,1/250
Momentum = 0,9
Nesterov = True

2.3 Voltar a configuração inicial e mudar o número de neurônios da camada LSTM

2.3.1 Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 16

4. Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 16:

RMSE	MSE	MAPE	MAE
371,92595	138328,9141	19,81251%	349,08027

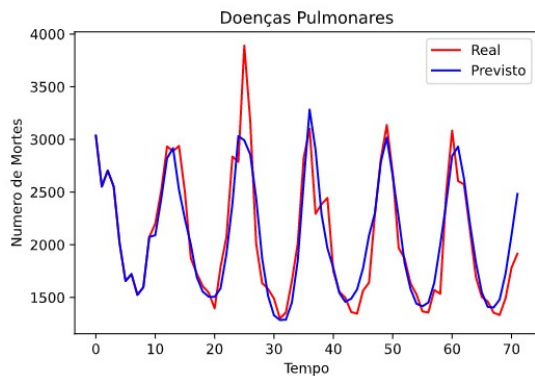


A precisão do modelo diminuiu, pois a métrica RMSE aumentou para, aproximadamente, 373. Isto significa que o modelo erra a previsão em 373 mortes e a métrica MAPE aumentou para, aproximadamente, 20%, isto é, a previsão está errada em 20%.

2.3.2 Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 48

5. Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 48:

RMSE	MSE	MAPE	MAE
239,98646	57593,50159	9,76142%	185,08326

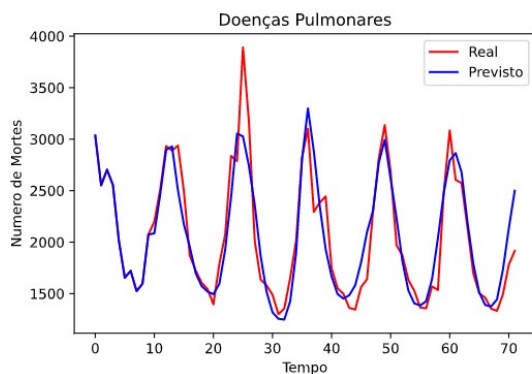


Os resultados do modelo ficaram muito próximos utilizando 100 e 48 neurônios, com incremento de somente 3% em relação ao resultado original. Mostrou-se a melhor configuração para a rede neural até esse momento.

2.4 Colocar uma segunda camada LSTM:

6. Colocar uma segunda camada LSTM:

RMSE	MSE	MAPE	MAE
242,59384	58851,77269	10,13818%	195,94057



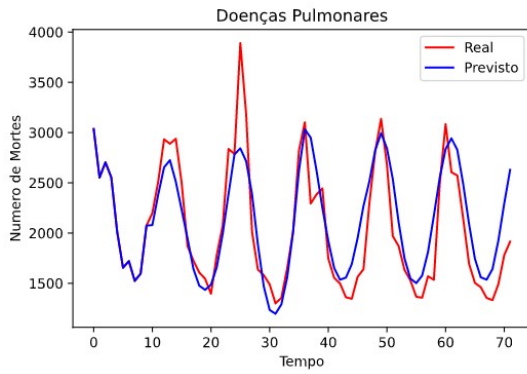
A adição de mais uma camada LSTM não melhorou o resultado do modelo, se aproximando muito do resultado da configuração original e aumentou o tempo de treinamento.

2.5 Mudar o número de épocas

2.5.1 Mudar o número de épocas para 100:

7. Mudar o número de épocas para 100:

RMSE	MSE	MAPE	MAE
362,76356	131597,3973	17,44883%	323,51646

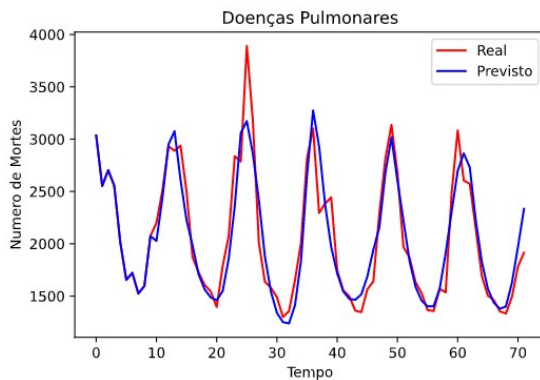


Mudar o número de épocas de 250 para 100 diminui a precisão do modelo, pois ficou claro que esse número é insuficiente para que o treinamento da rede neural seja completado.

2.5.2 Mudar o número de épocas para 500

8. Mudar o número de épocas para 500:

RMSE	MSE	MAPE	MAE
205,32314	42157,59364	7,91441%	162,19358

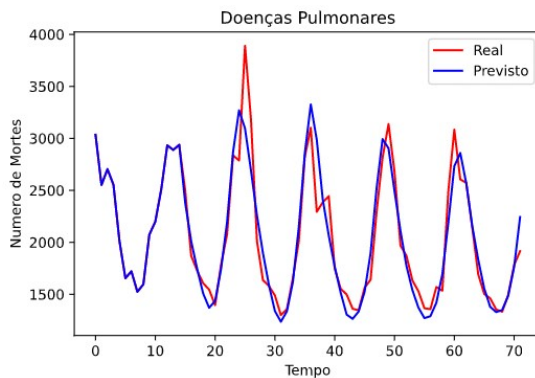


Com a mudança do número de épocas de 250 para 500, chegamos à configuração que tem a melhor precisão até o momento. diminui a precisão do modelo, pois ficou claro que esse número é insuficiente para que o treinamento da rede neural seja completado.

3 Com as mudanças feitas proponha sua configuração ideal

9. Janela 15 e output 1, 3 camadas LSTM e 500 épocas:

RMSE	MSE	MAPE	MAE
164,99173	27222,27048	5,10985%	108,22217



A configuração ideal foi encontrada com a criação de uma estrutura de dados com janela de input de 15 e output de 1 e acrescentando a terceira camada LSTM. O experimento mostrou que aumentar a janela de entrada pode beneficiar o desempenho do modelo, porém uma janela muito grande, pode potencialmente levar a um sobre-treino e degradar seu desempenho.

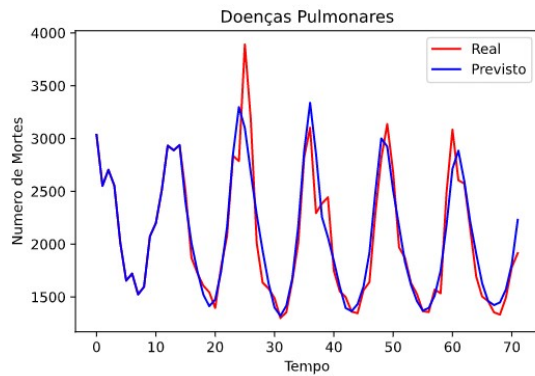
Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 15, 48)	9600
dropout (Dropout)	(None, 15, 48)	0
lstm_1 (LSTM)	(None, 80)	41280
dropout_1 (Dropout)	(None, 80)	0
dense (Dense)	(None, 1)	81
Total params: 50,961		
Trainable params: 50,961		
Non-trainable params: 0		

4 Comparação dos resultados utilizando GRU

10. Janela 15 e output 1, 3 camadas GRU e 500 épocas:

RMSE	MSE	MAPE	MAE
184,36781	33991,48958	7,08343%	140,66864



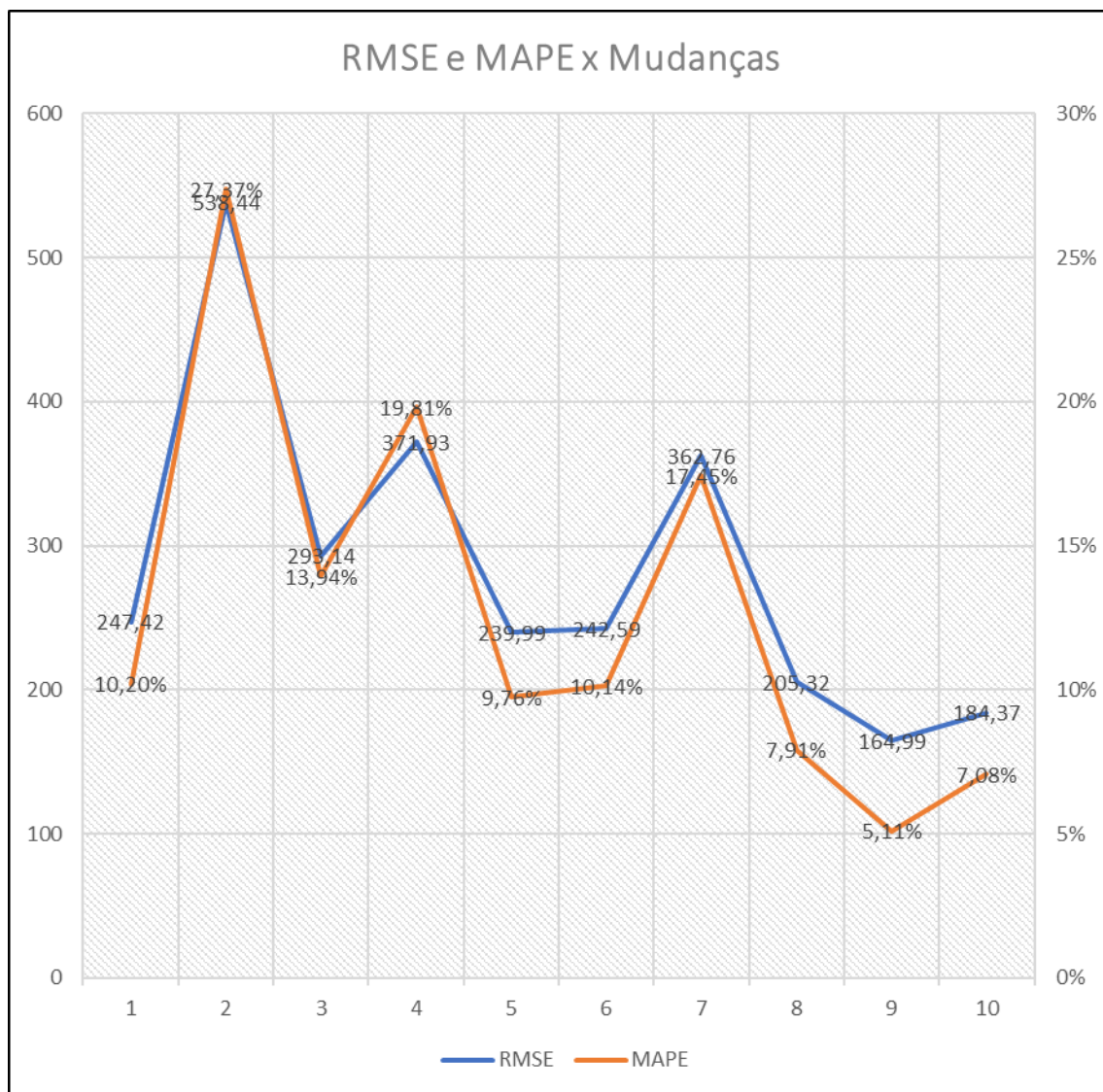
Substituímos as camadas LSTM por camadas GRU e o resultado apresentado ficou abaixo somente da configuração ideal, utilizando LSTM. A GRU tem resultados parecidos com os da LSTM, com tempo de processamento menor.

Model: "sequential"

Layer (type)	Output Shape	Param #
gru (GRU)	(None, 15, 48)	7344
dropout (Dropout)	(None, 15, 48)	0
gru_1 (GRU)	(None, 80)	31200
dropout_1 (Dropout)	(None, 80)	0
dense (Dense)	(None, 1)	81
Total params: 38,625		
Trainable params: 38,625		
Non-trainable params: 0		

5 Resultados alcançados em cada uma das Mudanças

1. Configuração Original			
RMSE	MSE	MAPE	MAE
247,41615	61214,75030	10,19675%	195,67470
2. Mudar o otimizador de adam para sgd dentro da configuração da rede:			
RMSE	MSE	MAPE	MAE
538,44489	289922,90234	27,37066%	478,54623
3. Mudar o otimizador de adam para sgd, utilizando Nesterov momentum e learning rate 0,1:			
RMSE	MSE	MAPE	MAE
293,13725	85929,44749	13,93915%	264,62549
4. Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 16:			
RMSE	MSE	MAPE	MAE
371,92595	138328,9141	19,81251%	349,08027
5. Voltar a configuração inicial (adam) e mudar o número de neurônios da camada LSTM para 48:			
RMSE	MSE	MAPE	MAE
239,98646	57593,50159	9,76142%	185,08326
6. Colocar uma segunda camada LSTM:			
RMSE	MSE	MAPE	MAE
242,59384	58851,77269	10,13818%	195,94057
7. Mudar o número de épocas para 100:			
RMSE	MSE	MAPE	MAE
362,76356	131597,3973	17,44883%	323,51646
8. Mudar o número de épocas para 500:			
RMSE	MSE	MAPE	MAE
205,32314	42157,59364	7,91441%	162,19358
9. Janela 15 e output 1, 2 camadas LSTM e 500 épocas:			
RMSE	MSE	MAPE	MAE
164,99173	27222,27048	5,10985%	108,22217
10. Janela 15 e output 1, 2 camadas GRU e 500 épocas:			
RMSE	MSE	MAPE	MAE
184,36781	33991,48958	7,08343%	140,66864



6 Conclusão

A partir da configuração inicial proposta no trabalho passamos pelas etapas de mudança dos otimizadores com o intuito de verificar se os resultados seriam mais assertivos. Introduzimos o SGD padrão e notamos, como já esperado, que os resultados apresentaram uma piora significativa, uma vez que esse otimizador não apresenta bom desempenho para treino de redes recorrentes como a LSTM.

Visto isso, retornamos com o otimizador Adam, efetuamos as modificações solicitadas e incluímos outras, de forma a alcançar melhores resultados. O processo foi apresentado na tabela do item 5, onde o melhor desempenho foi demonstrado no tópico 9.

Em síntese, as modificações foram:

- Janela de 15 com *output* 1,
- 2 camadas LSTM,
- 500 épocas de treinamento.

Resultando em um RMSE de 165 e MAPE 5%.

Adicionalmente, fizemos um teste substituindo a rede LSTM pela rede GRU. Apesar de alcançarmos bons resultados, a rede LSTM ainda apresentou o melhor desempenho.

Por fim, é importante ressaltar que sempre se faz necessário verificar os resultados dos indicadores, uma vez que os gráficos podem indicar uma melhora visual que, entretanto, na prática não é verificada nas métricas de desempenho.