

# Relatório - EP 1 - RNNs Bidirecionais, Overfitting, Underfitting

## MAC5725 - Linguística Computacional

### Introdução

Neste relatório, aplicaremos redes neurais recorrentes bidirecionais com arquiteturas LSTM (Long Short-Term Memory) e GRU (Gated Recurrent Unit) ao problema de análise de sentimentos. O corpus de avaliações da B2W será empregado como conjunto de dados, com atenção voltada para as colunas texto (`review_text`) e rótulo (`overall_rating`).

Além da implementação prática dessas arquiteturas, serão abordados desafios recorrentes no aprendizado de máquina, destacando-se o subajuste e o sobreajuste. Estratégias fundamentais, como a inclusão da camada Dropout durante o treinamento e a divisão do corpus em conjuntos de treinamento, validação e teste, serão exploradas como meios de enfrentar essas problemáticas.

### Configurações e hiperparâmetros utilizados

O pré-processamento dos dados é feito em R (`src/pre_processamento.R`). Do total de 132.373 avaliações, são filtradas 129.098 observações com algum texto de avaliação e notas entre 0 e 5. A divisão da base em treino, validação e teste foi feita com proporções de 65%, 10% e 25%, respectivamente, utilizando uma semente (923) e a função `initial_validation_split` do `rsample`.

Para a codificação dos textos, foi utilizada uma camada de embedding treinada com os próprios dados de treinamento.

Durante o treinamento dos modelos, foram escolhidos os hiperparâmetros `tammax = 100` e `batch_size = 256`, que demonstraram uma adaptação mais eficaz às metas estabelecidas. O hiperparâmetro `dropout` será avaliado a partir dos gráficos de validação gerados com três valores possíveis: 0, 25% e 50%.

## Validação

Como comentado anteriormente, vamos verificar como as medidas de perda (loss) e acurácia se comportam com os modelos com valores de **dropout** diferentes.

### Modelo com Encoder unidirecional

Warning: Removed 20 rows containing missing values (``geom_line()``).

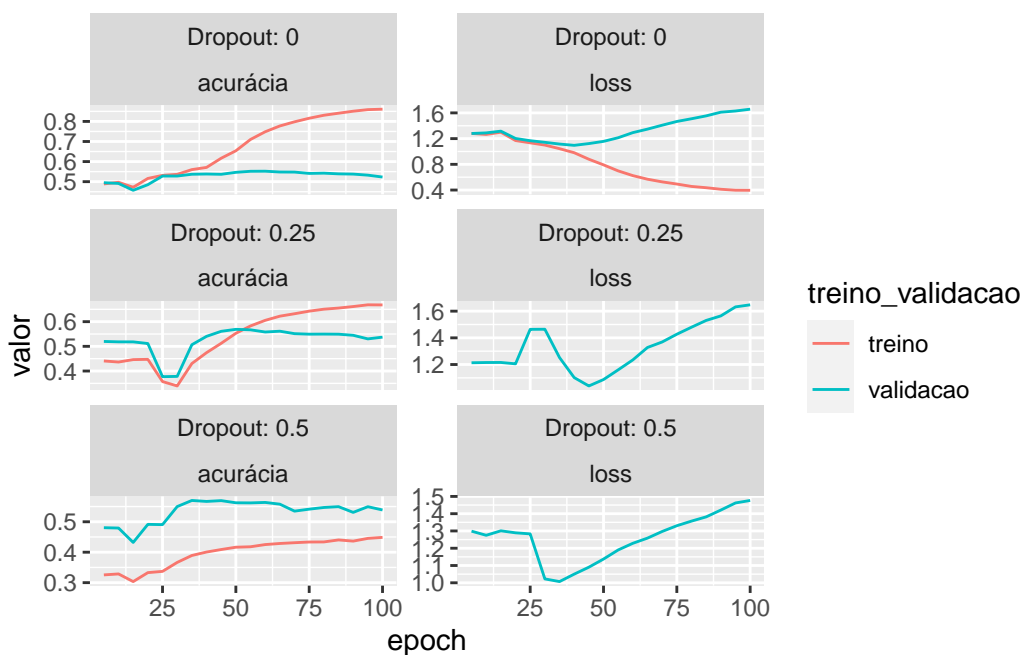


Figura 1: Métricas do modelo com Encoder unidirecional

A partir dos gráficos observados na Figura 1, foi estabelecido o valor de 50 épocas para o treinamento, a partir do qual a loss no conjunto de validação passa a aumentar e as curvas de acurácia passam a divergir significativamente no modelo com dropout de 25%.

### Modelo com Encoder bidirecional

Os modelos com Encoder bidirecional tiveram um comportamento muito diferente no treinamento quando comparado aos modelos anteriores.

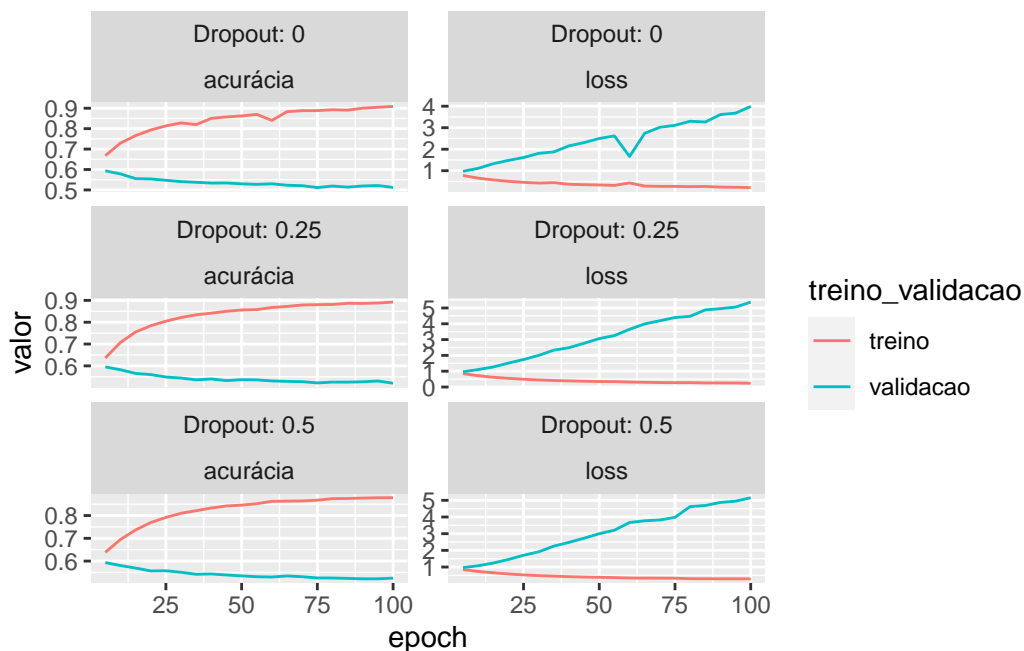


Figura 2: Métricas do modelo com Encoder unidirecional

Com certos hiperparâmetros, o treinamento retornou valores NaN para a loss, o que pode ser devido a uma taxa de aprendizado muito alta, instabilidade numérica nos gradientes, tamanhos de lote inadequados ou complexidade excessiva do modelo.

É interessante notar que com poucas épocas ele teve medidas de acurácia comparáveis às medidas dos modelos com Encoder unidirecional. Para efeitos de comparação entre os modelos, vamos escolher o valor de 20 épocas de treinamento do modelo.

## Teste

A Tabela 1 mostra as medidas de acurácia de teste dos seis modelos treinados.

Tabela 1: Acurácia de teste

Test Accuracy	Encoder	Dropout
0.5042293	unidirecional	0%
0.5638110	unidirecional	25%
0.5459024	unidirecional	50%
0.5442603	bidirecional	0%
0.5531216	bidirecional	25%

Test Accuracy	Encoder	Dropout
0.5781565	bidirecional	50%

A comparação dos resultados dos seis modelos revela insights interessantes sobre o impacto do tipo de encoder (unidirecional ou bidirecional) e a taxa de dropout na precisão do teste.

Inicialmente, observa-se que os modelos bidirecionais, em média, superaram seus equivalentes unidirecionais. Esse aumento na capacidade de capturar contextos em ambas as direções pode ter contribuído para o desempenho superior.

Além disso, a introdução de dropout teve um efeito variado nos resultados. Para os modelos unidirecionais, uma taxa de dropout de 25% pareceu beneficiar a precisão em comparação com o modelo sem dropout, enquanto uma taxa de 50% resultou em uma leve queda. Já nos modelos bidirecionais, a taxa de dropout de 50% se destacou, proporcionando a mais alta precisão entre todas as configurações. Esses resultados destacam a importância da escolha cuidadosa da arquitetura do modelo e da regularização para otimizar o desempenho em tarefas de predição de scores de avaliações.