

Relatório Aprendizado de Máquina Aula 9

Lucas Ribeiro da Silva - 2022055564

Universidade Federal de Minas Gerais

Belo Horizonte - Minas Gerais - Brasil

lucasrsilvak@ufmg.br

1 Introdução

Nesse relatório, procuraremos comparar o desempenho dos modelos de classificação Random Forest e Árvores de Decisão, de regressão Gradient Boosting e Árvores de Decisão e buscar compreender o impacto dos respectivos hiperparâmetros.

2 Modelos de Classificação

Neste primeiro teste, executaremos os métodos de classificação, utilizando 100 estimadores para comparar os métodos de Árvore de Decisão e Random Forest.

Tabela 1: Comparação dos Métodos de Classificação

Modelo	Métrica	Valor	Matriz de Confusão
Árvore de Decisão	Precisão (Classe 0)	0.74	$\begin{bmatrix} 32 & 17 \\ 11 & 30 \end{bmatrix}$
	Recall (Classe 0)	0.65	
	F1-Score (Classe 0)	0.70	
	Precisão (Classe 1)	0.64	
	Recall (Classe 1)	0.73	
	F1-Score (Classe 1)	0.68	
Random Forest	Precisão (Classe 0)	0.86	$\begin{bmatrix} 43 & 6 \\ 7 & 34 \end{bmatrix}$
	Recall (Classe 0)	0.88	
	F1-Score (Classe 0)	0.87	
	Precisão (Classe 1)	0.85	
	Recall (Classe 1)	0.83	
	F1-Score (Classe 1)	0.84	

2.1 Variando Estimadores

Neste segundo teste, variaremos o número de estimadores e observaremos a resposta do classificador Random Forest.

Nº de Estimadores	Métrica	Valor	Matriz de Confusão
50	Precisão (Classe 0)	0.86	$\begin{bmatrix} 43 & 6 \\ 7 & 34 \end{bmatrix}$
	Recall (Classe 0)	0.88	
	F1-Score (Classe 0)	0.87	
	Precisão (Classe 1)	0.85	
	Recall (Classe 1)	0.83	
	F1-Score (Classe 1)	0.84	
200	Precisão (Classe 0)	0.84	$\begin{bmatrix} 42 & 7 \\ 8 & 33 \end{bmatrix}$
	Recall (Classe 0)	0.86	
	F1-Score (Classe 0)	0.85	
	Precisão (Classe 1)	0.82	
	Recall (Classe 1)	0.80	
	F1-Score (Classe 1)	0.81	

Não foram observadas grandes variações nos resultados variando o número de estimadores, isso pode significar que 50 estimadores já eram suficientes para a complexidade do problema sugerido.

2.2 Conclusão da comparação

Podemos perceber que o Random Forest tem uma resposta consideravelmente melhor que uma única Árvore de Decisão, isso ocorre pois combina diversas árvores de decisão aleatórias, permitindo maior generalização no conjunto de dados e reduzindo a chance de overfit

3 Modelos de Regressão

Neste segundo teste, executaremos os métodos de regressão, utilizando 100 estimadores para comparar os métodos de Random Forest e Gradient Boosting, variaremos algumas das informações e parâmetros e faremos as observações necessárias.

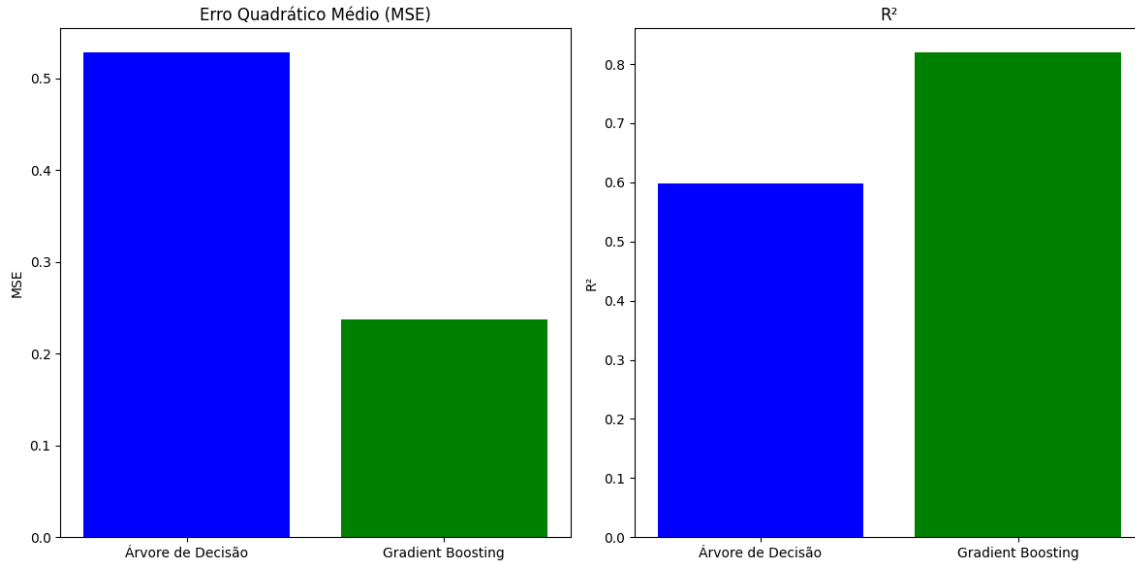


Figura 1: Modelos de Regressão

3.1 Variação de Estimadores

Para esse teste, variaremos o número de estimadores no Gradient Booster.

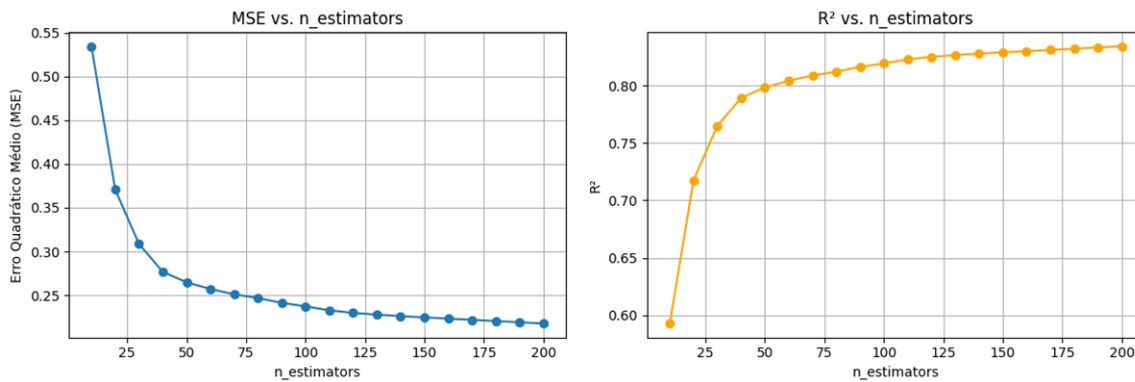


Figura 2: Variação de Estimadores

Pelo gráfico, é possível concluir que o modelo tende a ter maior precisão e acurácia conforme o número de estimadores aumenta.

3.2 Variação da Taxa de Aprendizado

Para esse teste, variaremos a taxa de aprendizado no Gradient Booster.

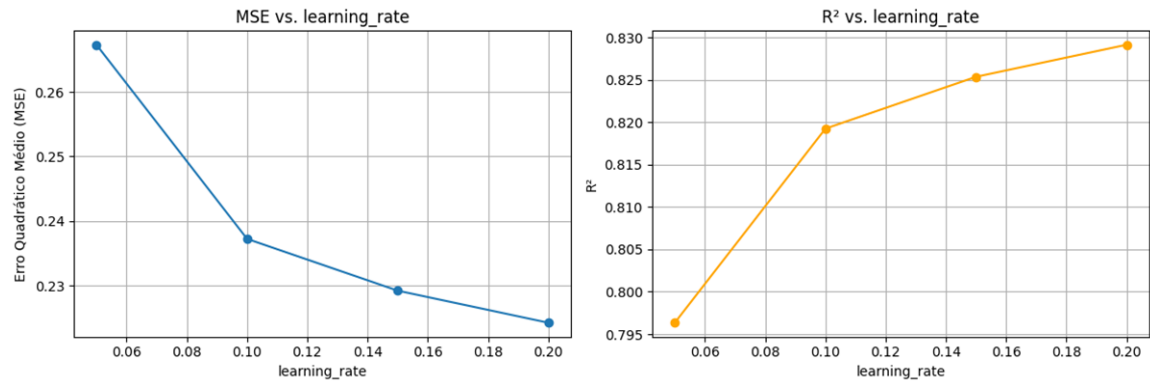


Figura 3: Variação da Taxa de Aprendizado

Pelo gráfico, é possível concluir que o modelo tende a ter maior precisão e acurácia conforme a taxa de aprendizado aumenta.

3.3 Variação da Profundidade Máxima

Para esse teste, variaremos a profundidade máxima no Gradient Booster.

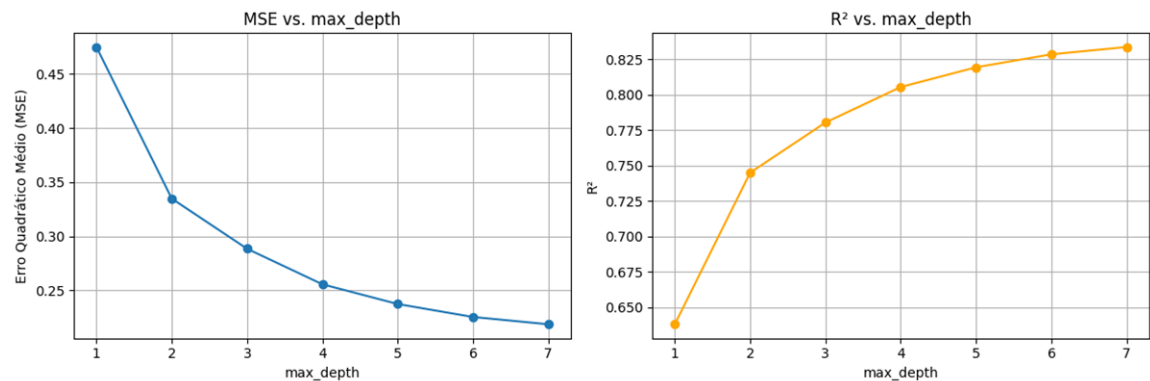


Figura 4: Variação da Profundidade Máxima

Pelo gráfico, é possível concluir que o modelo tende a ter maior precisão e acurácia conforme a profundidade máxima aumenta.

4 Conclusões

4.1 Padronização

Os métodos baseados em árvores, como Árvores de Decisão, Random Forest e Gradient Boosting não precisam de padronização dos dados pois os limiares são calculados localmente, dividindo com base em comparações diretas dos valores das variáveis.

4.2 Robustez

Os métodos mencionados são ainda robustos, pois agrega diferentes árvores independentes por meio da tática de *bootstrap* além da utilização de seleção aleatória, que reduz o *overfitting* consideravelmente.