

Exercício para sala de aula - Árvores de decisão e Ensembles

João Paulo Mota

August 21, 2024

1 Roteiro do exercício

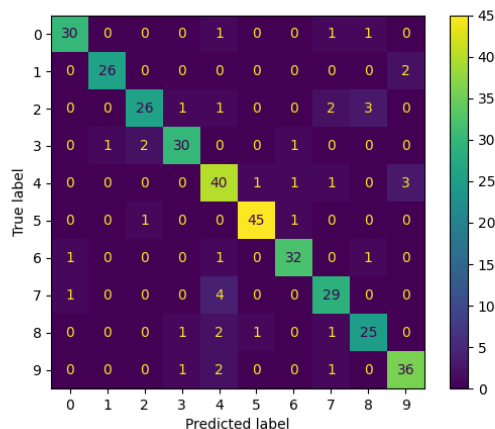
1.1 Exercício 1 - Treinamento do modelo baseado em árvore de decisão

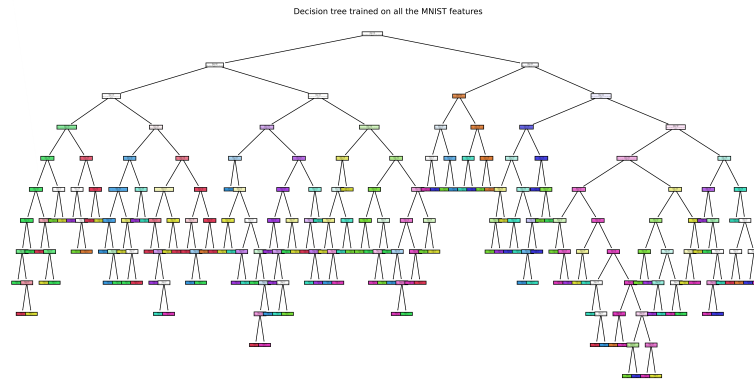
O treinamento do modelo de árvore de decisão foi feito utilizando a biblioteca GridSearchCV para busca de melhores parâmetros. Sabemos que a qualidade dos candidatos a split de um nó e, por consequência, a escolha feita pelo modelo em treinamento como dividir o nó, é feita analisando funções de impureza ou de perda. Aqui analisamos log_loss, gini impurity e entropy. Os resultados foram:

- criterion=log_loss, precision=0.8861, training_time=0.1876, max_depth=None
- *criterion=entropy, precision=0.8861, training_time=0.1848, max_depth=None
- criterion=gini, precision=0.8416, training_time=0.0960, max_depth=None

Note que escolhemos não setar uma profundidade máxima pois o dataset é relativamente pequeno e é vantajoso que o modelo continue fazendo splits até que todas as folhas sejam puras (pertencam a uma só classe).

Abaixo vemos a matrix de confusão e a árvore resultando para os parâmetros escolhidos (linha marcada com *)





1.2 Exercício 2 - Avaliação dos ganhos com a utilização de modelos Ensemble

Vemos um ganho substancial de performance utilizando modelos Ensemble.

Abaixo vemos dados de análise dos modelos Ensemble utilizados:

Random Forest:

- criterion=log_loss, precision=0.9701, training_time=0.1876, max_depth=None
- *criterion=entropy, precision=0.9722, training_time=0.1848, max_depth=None
- criterion=gini, precision=0.9722, training_time=0.0960, max_depth=None

XGBoost:

- max_depth=5, training_time=1.5073583126068115, precision=0.9666666666666667
- max_depth=10, training_time=1.6833817958831787, precision=0.9694444444444444
- max_depth=100, training_time=1.5776081085205078, precision=0.9694444444444444
- max_depth=1000, training_time=1.555584192276001, precision=0.9694444444444444

1.3 Exercício 3 - Visualização da árvore de decisão e Medida de Impureza

Olhar itens anteriores.

1.4 Exercícios 4, 5 e 6

Ver código python que acompanha esse relatório e arquivo README