

Faculdade de Informática e Administração Paulista – FIAP
Campus Aclimação
Turma 1TDSR
Artificial Intelligence and ChatBot

CP5 - Machine Learning - Regressão Linear

Gabriel Nogueira Peixoto

12 de outubro de 2025.

São Paulo.

QUESTÕES AVALIATIVAS:

1 - Qual a importância de separar as bases em treino e teste?

Precisamos separar as bases em treino, que serve para de fato ser feito o treinamento com a maior parte da base de dados e também precisamos separar em teste, para avaliarmos se de fato o modelo conseguiu fazer previsões próximas da realidade com dados que o modelo nunca viu antes. Fazer essa separação nos permite fazer os ajustes necessários para que o modelo aprenda com os padrões dos dados e possamos detectar o overfitting.

2 - Do ponto de vista prático, quais são as vantagens de utilizarmos o sklearn?

As principais vantagens de utilizar o sklearn se dá pela sua biblioteca de classes extensas, que já contam com os principais tipos de algoritmo e ferramentas que nos auxiliam no treinamento de modelos, como algoritmo de Regressão Linear, Decision Tree Regressor e MSE para avaliar o modelo. Também podemos utilizar o sklearn, para criar um fluxo de trabalho mais limpo, através dos pipelines de pré-processamento que são disponibilizados.

3 - Qual dos modelos apresentados previu os valores médios das casas nos distritos da Califórnia: Regressão Linear ou Decision Tree Regressor? Justifique sua resposta.

Ambos os modelos conseguiram fazer a previsão dos valores médios das casas. O que aconteceu, é que utilizando o modelo de regressão linear, tivemos um **MAPE(Mean Absolute Percentage Error)** de 28%, o que quer dizer que esse modelo teve um overfitting e não desempenhou bem com dados reais (**também teve um erro de margem de 69 mil dólares**).

Já o modelo de Decision Tree Regressor, conseguiu ter um resultado superior, e teve um **MAPE** de 18%. Ou seja, conseguiu prever e desempenhar melhor com os dados teste, apenas mudando o modelo de treinamento.

4 - Cite um outro exemplo prático e contextualizado onde a regressão linear seria uma boa escolha para fazer previsão. Justifique sua resposta.

Um exemplo prático que poderíamos utilizar um algoritmo de regressão linear é no crescimento de plantas em jardinagem. Podemos estimar a altura/tamanho de uma planta ao longo do tempo (**label**) com base na relação linear com **features** como quantidade de água, horas de exposição a luz e fertilização.