

# Aula Prática 06

Todos os exercícios a seguir devem ser realizados usando a base de dados *Housing Prices* ([Housing Prices Dataset | Kaggle](#)). Não podem ser utilizadas bibliotecas além das pandas e numpy. A variável *price* é a variável alvo e as demais variáveis são as preditivas.

1. Implemente o One-Hot-Encoding para as variáveis categóricas.
2. Separe o conjunto de dados em treino e teste.
3. a) Implemente a regressão linear ao conjunto de dados usando o método dos mínimos quadrados (OLS).  
b) Avalie o desempenho do modelo usando as métricas MSE e  $R^2$  (você deve implementar essas métricas)
4. a) Implemente a regressão linear ao conjunto de dados usando o método do gradiente descendente (GD). Plote a curva de aprendizagem.  
b) Normalize as variáveis preditivas usando o método min-max e a variável alvo usando o método z-score. Repita o procedimento do item A.  
c) Avalie o desempenho do modelo aplicado ao conjunto de dados normalizados, modelo usando as métricas MSE e  $R^2$ .
5. Repita os procedimentos realizados no item 4, usando o método SGD (gradiente descendente estocástico) ao invés do GD.
6. Treine 10 vezes o algoritmo GD, calculando o MSE e o  $R^2$  sobre o conjunto de teste após cada iteração. Faça o mesmo com o algoritmo SGD. Faça um gráfico dos valores do MSE e  $R^2$  para ambos os algoritmos. Compare esses valores com os obtidos com o método OLS. Por que esses valores variam tanto em cada treinamento? Você consegue propor uma mudança nesses algoritmos que melhore o desempenho?