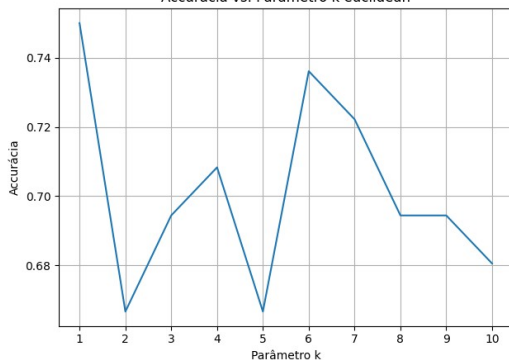


Relatório da atividade sobre KNN de Aprendizado de Máquina Supervisionado

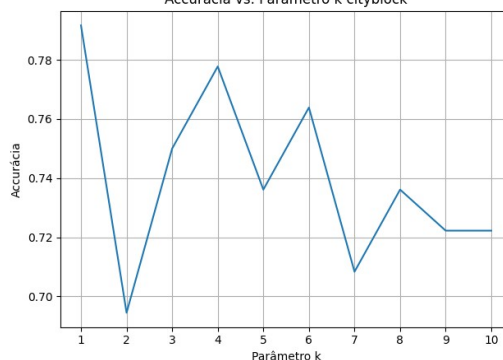
Vinicius Cordeiro Fonseca
20250082351
UFRN/IMD
07/08/2025

- 1) Obtivemos a base de dados “wine” disponível no pacote de datasets do scikit-learn e visualizamos suas 7 primeiras linhas.
- 2) Observamos as estatísticas descritivas da base de dados wine.
- 3) Para separar os dados em 2 subconjuntos (treinamento e teste), primeiro precisamos carregar os atributos como arrays do numpy. Então dividimos a base de dados em treinamento e teste.
- 4) Calculando a acurácia com valores de K observa-se que chebyshev e cityblock possuem maior acurácia com K igual a 4 enquanto a distância euclidiana tem maior acurácia com K igual a 6; percebemos também que chebyshev e cityblock obtiveram acurácia mais alta que quando usamos a distância euclidiana com dados não normalizados.
- 5) A seguir apresentamos os gráficos da variação da acurácia com o parâmetro K para cada distância utilizada:

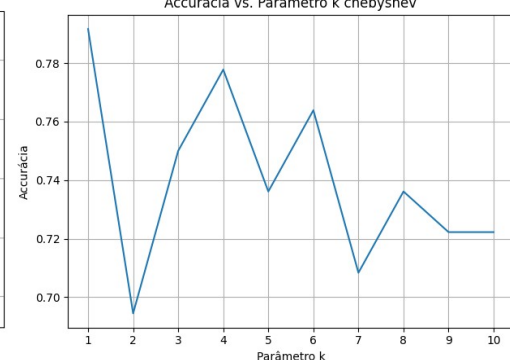
Accurácia vs. Parâmetro k euclidean



Accurácia vs. Parâmetro k cityblock

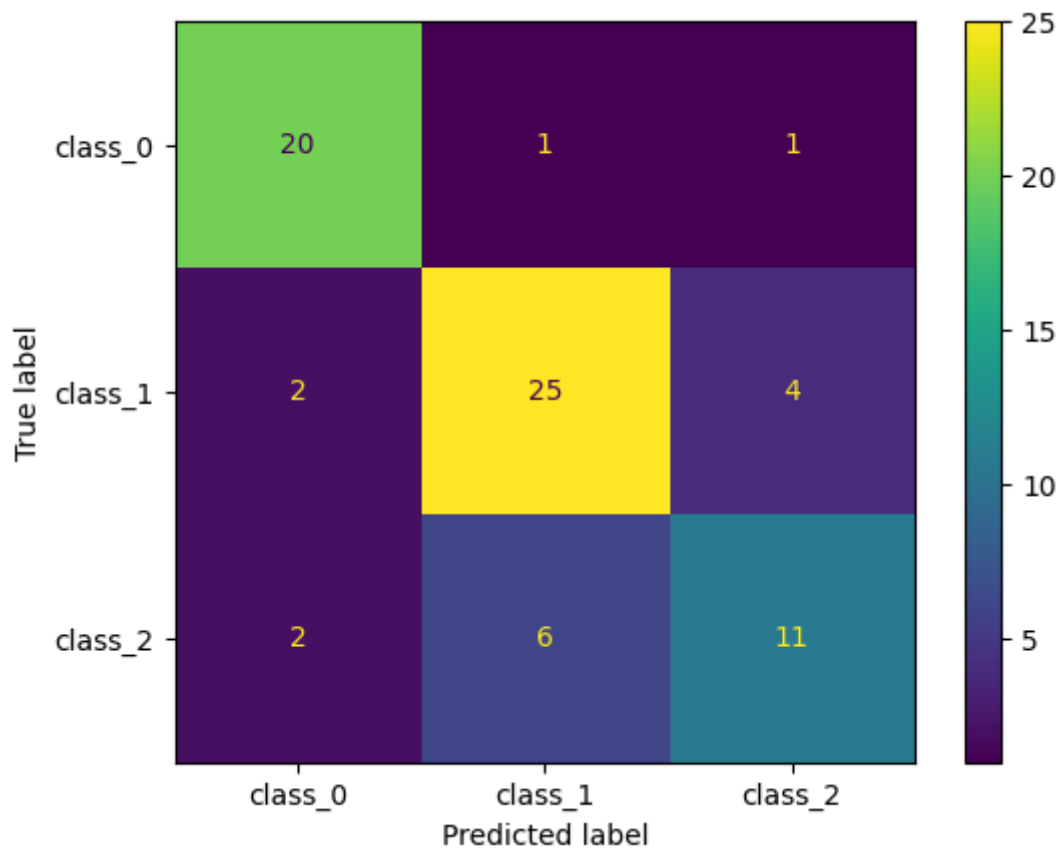


Accurácia vs. Parâmetro k chebyshev



6) Após determinar o melhor número de K pegamos a acurácia de cada um dos modelos e decidimos selecionar o modelo com cityblock: (euclidean, K=6, acurácia = 0.73); (cityblock. K=4, acurácia = 0.77); (chebyshev, K = 4, acurácia = 0.72)

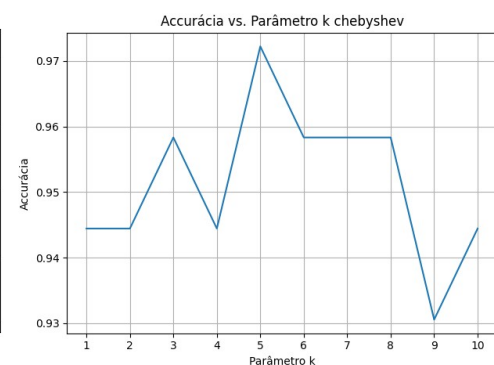
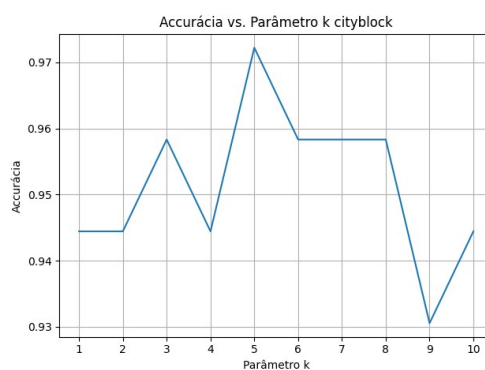
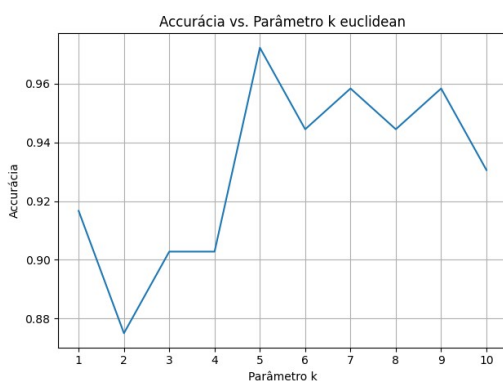
7) O modelo com cityblock foi treinado, então calculada sua matriz de confusão e gerada sua figura:



8) A seguir repetimos os itens 4-7 com os atributos normalizados:

8.4) Nota-se que a acurácia melhorou significativamente em todas as distâncias utilizadas, e a melhor acurácia em todos foi com K igual a 5.

8.5) A seguir apresentamos os gráficos da variação da acurácia com o parâmetro K para cada distância utilizada



8.6) Após descobrir o melhor valor de K, igual a 5 para todos os modelos, pegamos a acurácia de cada um dos modelos e decidimos selecionar o modelo com euclidean: (euclidean, acurácia = 0.97); (cityblock. acurácia = 0.97); (chebyshev, acurácia = 0.88)

8.7) O modelo euclidean foi treinado, então calculamos sua matriz de confusão e geramos sua figura:

