

Universidade Federal de Pernambuco
Departamento de Eletrônica e Sistemas
Lista Machine Learning

Alunos: Adriel Filipe, Pedro Peixoto, Gabriel Garcia, Robson Lima, Mateus Henrique
Curso: Engenharia Eletrônica

1 Questão 2

Os códigos estão disponíveis em: https://github.com/yriuss/MagicLearning/tree/master/questao2_Farias

1.1 letra a

Para rodar o código, o Pandas e sklearn são necessários, embora não tenham sido usados para o método em si. Quando baixar os códigos do *GitHub*, basta rodar *questao2a.py*.

O KNN foi criado como uma classe. O método “fit” simplesmente pega parte dos dados (variável X) e os targets (variável y) para servir como treino. O método externo “predict”, chama o método interno para cada ponto com dimensão 13 (número de atributos para cada ponto) dentro dos dados X. Na função interna, cada ponto x tem sua distância com cada um dos pontos pegos no método “fit” calculada. Assim, organizamos as k menores distâncias e pegamos o target das mais comuns, como o target escolhido. Plotamos a acurácia para k indo de 1 até 51.

1.2 letra b

Para rodar o código, o Pandas, seaborn e sklearn são necessários, embora não tenham sido usados para o método em si. Quando baixar os códigos do *GitHub*, basta rodar *questao2b.py*.

Para inicialização, alguns pontos convenientes dentro dos dados foram escolhidos como centroides iniciais. Foram escolhidos 3 centroides, dado que são 3 *targets*. No código desenvolvido, foram rodadas 50 iterações, podendo o número de iterações ser alterado sem problemas. O código basicamente calcula as distâncias dos centroides dos dados, escolhendo sempre os centroides mais próximos para cada dado, posteriormente atualizando os centroides e continuando o *looping* até sua convergência ($sum(distances) == 0$).

As taxas de acerto são printadas e são de 77.97%, 70.42% e 60.42%. No final, os dados clusterizados são plotados par a par em todas as 13 dimensões, o arquivo do plot dos dados está disponível em “Plotdados.png”.