

Lista 02

Exercícios sobre Aprendizado Baseado em Instâncias

Prof. Marcelo Keese Albertini
Universidade Federal de Uberlândia

Exercício 1

Em R, sem usar implementação de pacote pronto para o k -NN, escreva um código para implementar o método de aprendizado 1NN para classificação.

Exercício 2

Em R, escreva um código usando `tidymodels`, para implementar o método de aprendizado k -NN usando distância de Manhattan.

Exercício 3

Se o atributo-alvo de um problema de aprendizado de máquina for contínuo, como você faria para usar aprendizado baseado em instância para esse problema?

Exercício 4

Se todas as características de observações em um dataset for categórica, como você faria para usar aprendizado baseado em instância para esse dataset?

Exercício 5

Discuta as vantagens e desvantagens do aprendizado baseado em instâncias?

Exercício 6

Como deve-se usar a distância dos vizinhos mais próximos para realizar o aprendizado baseado em instâncias?

Exercício 7

Discorra sobre a maldição da dimensionalidade.

Exercício 8

Escreva o algoritmo de seleção direta e explique para qual situação ele é útil.

Exercício 9

Escreva o algoritmo de eliminação retroativa e explique para qual situação ele é útil.

Exercício 10

O que é o algoritmo k -NN editado? Em quais situações devemos usá-lo? Cite pelo menos 2 situações diferentes.

Exercício 11

Escreva um algoritmo de validação cruzada para a escolha do parâmetro k para o treino do k -NN (sem seleção de atributos e sem edição).

Exercício 12

Quais são diferenças entre aprendizado lazy (preguiçoso) e greedy (impaciente/ganancioso)?

Exercício 13

Qual é a relação entre redes de base radial e o k -NN?

Exercício 14

Porque pode ser importante colocar os atributos numéricos de um dataset em uma mesma escala? Escreva um código em R, usando `tidymodels` para fazer isso.

Exercício 15

Provou-se que $\lim_{n \rightarrow \infty, k \rightarrow \infty, k/n \rightarrow 0} \epsilon_{k\text{-NN}} = \epsilon^*$, onde $\epsilon_{k\text{-NN}}$ representa o erro obtido pelo k -NN e ϵ^* é o menor erro teórico para uma distribuição de dados. Explique o que representa esse limite. Justique, informalmente, porque é necessário que $k/n \rightarrow 0$.