Aula Prática 4

Murielly Oliveira Nascimento – 11921BSI222

Tema: Classificação – Algoritmo KNN

1. Execute o algoritmo KNN na seguinte base de dados para classificar a seguinte instância. Mostre o passo a passo da execução.

#11 A1=5 A2=7

Considere K=1,3,5 e 7.

|  |  |  |  |
| --- | --- | --- | --- |
| # | A1 | A2 | Classe |
| 1 | 0,5 | 1 | 2 |
| 2 | 2,9 | 1,9 | 2 |
| 3 | 1,2 | 3,1 | 2 |
| 4 | 0,8 | 4,7 | 2 |
| 5 | 2,7 | 5,4 | 2 |
| 6 | 8,1 | 4,7 | 1 |
| 7 | 8,3 | 6,6 | 1 |
| 8 | 6,3 | 6,7 | 1 |
| 9 | 8 | 9,1 | 1 |
| 10 | 5,4 | 8,4 | 1 |

|  |  |
| --- | --- |
| Pontos | Distância Euclidiana em Relação ao P11 |
| 1 | 7,5 |
| 2 | 5,515432893 |
| 3 | 5,445181356 |
| 4 | 4,788527958 |
| 5 | 2,801785145 |
| 6 | 3,860051813 |
| 7 | 3,324154028 |
| 8 | 1,334166406 |
| 9 | 3,661966685 |
| 10 | 1,456021978 |

|  |  |  |
| --- | --- | --- |
| Valores de K | Pontos mais próximos | Classe |
| K=1 | P8 | 1 |
| K=3 | P8, P10, P5 | 1 |
| K=5 | P8, P10, P5, P7, P9 | 1 |
| K=7 | P8, P10, P5, P7, P9, P6, P4 | 1 |

1. Implementação sobre técnicas de Classificação
   1. Busque na Internet uma implementação do algoritmo KNN feita na sua linguagem de preferência.

<https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>

* 1. Discuta cada linha do código identificando qual ação está sendo realizada.
  2. Qual o formato dos dados da entrada? Qual distância foi utilizada? O algoritmo implementado faz algum pré-processamento na base?

Os dados são numéricos com três atributos. A distância usada pelo algoritmo é a Euclidiana. Nenhum pré-processamento foi feito.

* 1. Qual a saída do programa?

O programa mostra a saída esperada com base nos dados de treino e a previsão feita pelo algoritmo KNN.

* 1. Você acha que esse código ficou interessante? Alguma coisa poderia ser melhorada? Você faria uma implementação diferente?

O código é simples, eficiente e bem-organizado. Contudo, ele não faz nenhum pré-processamento. Dependendo da base de dados seria necessário tirar uma amostra ou normalizar os atributos.

* 1. Tente executar o código usando uma base de dados do UCI e variando valores de K.