**Aprendizado de Máquina na Prática**

**Implementação do *KNN – K Neareast Neighbor* em Python**

**para classificação da IRIS**

**Dionisio Gause Junior**

*Centro Universitário Leonardo da Vinci – UNIASSELVI*

*dionisio@informatyk.com.br; dgausejr@gmail.com*

O Aprendizado de Máquina ou “*Machine Learning”,* têm se destacado nas mais variadas áreas, ou seja, treinar o computador para responder rapidamente a uma questão, com base em dados fornecidos, comparando-os, e desta forma também “aprendendo” conforme mais dados são acrescentados, define-se como Método Supervisionado.

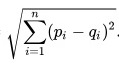
Ao abordar esta forma, escolhemos um dos métodos para demonstrar o quanto a área de Inteligência Artificial está se tornando cada vez mais essencial ao nosso dia a dia.

A proposição é a de criar no Python um algoritmo de *Machine Learning ML*, utilizando o método Supervisionado de Classificação - KNN (*K Neareast Neighbor*) ou K Vizinhos mais Próximos.

Para tal intento abordamos um pouco dos conceitos utilizados para este desenvolvimento.

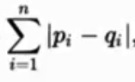
“In the absence of prior knowledge, most kNN classifiers use simple Euclidean distances to measure the dissimilarities between examples represented as vector inputs.”[[1]](#footnote-1)

“Na ausência de conhecimento prévio, a maioria dos classificadores de kNN usa distâncias euclidianas simples para medir as diferenças entre exemplos representados como entradas de vetores.”



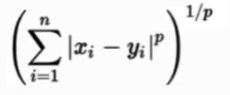
*Fig1. Fórmula Euclidiana[[2]](#footnote-2)*

Outra métrica para a classificação é a de Manhattan.



*Fig2. Fórmula de Manhattan[[3]](#footnote-3)*

A apresentação destas duas métricas são importantes, pois, são delas que a forma generalista de Minkowski se baseia, ou seja, ao informarmos o valor de *p = 2* utilizaremos a forma Euclidiana e/ou ao informar o valor de *p = 1* utilizamos a métrica de Manhattan.



*Fig3. Fórmula Minkowski[[4]](#footnote-4)*

Esta abordagem se faz necessária, pois ao desenvolvermos o algoritmo proposto, utilizamos pacotes ou *Packages* necessários para a codificação na linguagem Python, onde as funções utilizadas nos ajudarão a chegar às soluções propostas.

Para configurar o ambiente de desenvolvimento utilizou-se a versão do *Python* 3.7.2 e a IDE o *JetBrains PyCharm Community Edition 2018.2.3 x64 e* as bibliotecas: *Numpy*, *Pandas* e *Sklearn*.

Obs.:

Atributos:

Comprimento da sépala,

Largura da sépala,

Comprimento da pétala e

Largura da pétala

- Rótulo

“*... sempre que importar o Pandas, utilize a regra de convenção. Isso faz com que pessoas que lerem seu código no futuro — incluindo você mesmo — possa identificar a biblioteca mais facilmente. Por regra, Pandas deve ser importado sob o nome de pd, assim: import pandas as pd*”[[5]](#footnote-5)

***Referências:***

<http://papers.nips.cc/paper/2795-distance-metric-learning-for-large-margin-nearest-neighbor-classification.pdf acessado em 00/00/0000 00:00h>

<https://paulovasconcellos.com.br/como-criar-seu-primeiro-aplicativo-de-machine-learning-7b6af291ba11 - acessado em 00/00/0000 00:00h>

<https://www.youtube.com/watch?v=zvmbB3315Ko acessado em 19/09/2018 10:00h>

1. <http://papers.nips.cc/paper/2795-distance-metric-learning-for-large-margin-nearest-neighbor-classification.pdf> [↑](#footnote-ref-1)
2. <https://www.youtube.com/watch?v=zvmbB3315Ko acessado em 19/09/2018 10:00h> [↑](#footnote-ref-2)
3. <https://www.youtube.com/watch?v=zvmbB3315Ko acessado em 19/09/2018 10:00h> [↑](#footnote-ref-3)
4. <https://www.youtube.com/watch?v=zvmbB3315Ko acessado em 19/09/2018 10:00h> [↑](#footnote-ref-4)
5. <https://paulovasconcellos.com.br/28-comandos-%C3%BAteis-de-pandas-que-talvez-voc%C3%AA-n%C3%A3o-conhe%C3%A7a-6ab64beefa93 – acessado em 20/09/2018 19:19h> [↑](#footnote-ref-5)