**Universidade Federal do Agreste de Pernambuco**

**Bacharelado em Ciências da Computação**

**Disciplina: Reconhecimento de Padrões.**

**Professor: Tiago B. A. de Carvalho.**

**Alunos: David Brito e Laisy Ferreira**

**Aprendizagem de Máquina:**

**k-NN**

**1**. (20 pontos) Descreva um problema de classificação para o qual seria adequado utilizar o k-NN e descreva um problema de classificação para o qual não seria adequado utilizar este classi cador. Justifique suas escolhas baseado nas vantagens e desvantagens do k-NN. Mostre pelo menos duas vantagens e duas desvantagens para cada exemplo.

Um problema de classificação que o k-NN poderia ajudar a resolver é o de classificar a cor de um determinado objeto, como uma bolinha por exemplo. A partir de uma determinada distância (k) o algoritmo é capaz de determinar se um objeto ainda não classificado faz parte de uma classe específica (uma cor, neste caso), ao comparar os atributos deste objeto com os atributos de outros objetos daquela determinada classe. O objeto será classificado como a classe cujos atributos mais se assemelham aos seus.

Como este exemplo pode ser resolvido utilizando apenas um cálculo de distância (seja ela euclidiana, ponderada, etc), o algoritmo k-NN funciona muito bem para este exemplo.

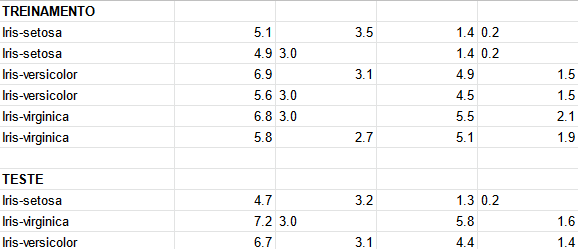
Entretanto, o k-NN não funciona muito bem com um conjunto de dados muito grande. Como o k-NN é um algoritmo baseado em distância, o custo de calcular a distância entre um ponto existente e um novo ponto é muito alto o que pode acabar afetando bastante a performance do algoritmo.

Além disso, se os dados já classificados estiverem a uma grande distância do objeto que está sendo determinado atualmente, fica difícil de garantir que aquele objeto faz parte de um dos conjuntos classificados.

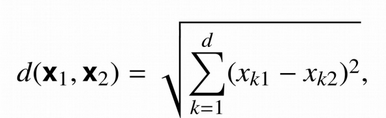
Já em um problema de reconhecimento facial em fotos (como o algoritmo que o Google Photos usa para classificar se fotos diferentes são de uma mesma pessoa), o k-NN não seria muito adequado, já que a base de dados é muito grande, com inúmeros atributos de classificação, o que tornaria a classificação muito custosa e imprecisa, já que este problema não tem relação com distância/proximidade.

**2**. (25 pontos) Utilizando a base de dados archive.ics.uci.edu/ml/datasets/iris:

(a) Selecione os três exemplos aleatórios de cada classe e construa a matriz de distância entre colocando um exemplo de cada classe como elemento de conjunto de teste e os outros 6 como conjunto de treinamento.



(b) Utilizando a matriz de distância explique a classificação dos exemplos de teste utilizando 1-NN.



* TESTE 1: Iris-setosa (treinamento) X Iris-setosa (teste)

X1 = (5.1 , 3.5, 1.4, 0.2) Y1 = (4.7, 3.2, 1.3, 0.2)

√(5.1 - 4.7)² + (3.5 - 3.2)² + (1.4 - 1.3)² + (0.2 - 0.2)²

√(0.4)² + (0.3)² + (0.1)² + (0)²

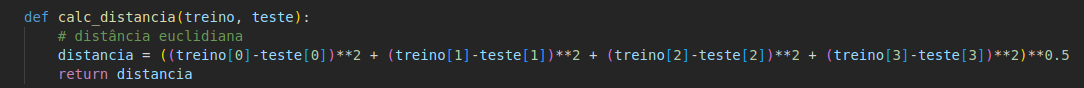
√(0.4)² + (0.3)² + (0.1)² + (0)²

√0.8 + 0.9 + 0.1

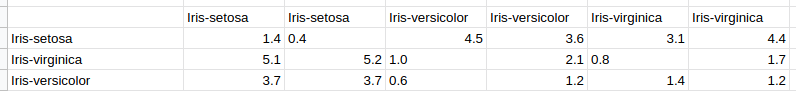
√1.8

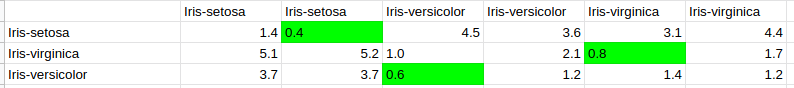
≃1.4

Criamos uma função para fazer todos os cálculos:

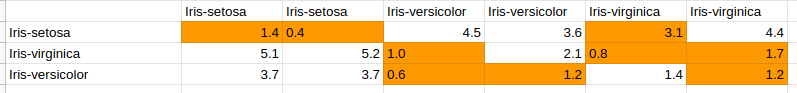


MATRIZ DE DISTÂNCIAS

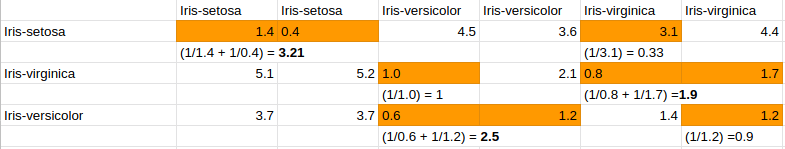




(c) Utilizando a matriz de distância explique a classificação dos exemplos de teste utilizando 3-NN sem peso.



(d) Utilizando a matriz de distância explique a classificação dos exemplos de teste utilizando 3-NN com peso.



(e) Selecione duas características da base Iris construa um diagrama de dispersão colocando símbolos ou cores distintas para cada classe.

Dica: esta questão pode ser resolvida inteiramente no Excel ou no LibreOffice Calc.

**3**. (15 pontos) Utilize o classi cador pelo vizinho mais próximo utilizando distância euclidiana. Avalie este classi cador utilizando metade dos exemplos de cada classe da base Iris como conjunto de teste e o restante como conjunto de treinamento. Utilize uma biblioteca para o classi cador 1-NN. Dica: você pode utilizar o sklearn scikit-learn.org/stable/modules/ generated/sklearn.neighbors.KNeighborsClassifier.html.

**4**. (20 pontos) Faça o mesmo da questão anterior sem utilizar biblioteca, você deve implementar o 1-NN com distância euclidiana.

**5**. (20 pontos) Utilize os classi cadores 7-NN com e 7-NN sem peso e avalie os classi cadores utilizando metade dos exemplos de cada classe da base Speaker Accent Recognition como conjunto de teste e a outra metade como conjunto de treinamento. Base: archive.ics.uci. edu/ml/datasets/Wine, arquivo accent-mfcc-data-1.csv. Dica: você pode utilizar o skle arn archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition.