

Disciplina: Aprendizagem de Máquina.
Professor: Tiago B. A. de Carvalho.

Exercícios da Semana 02 – 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 classificador. Justifique suas escolhas baseado nas vantagens e desvantagens do k-NN. Mostre pelo menos duas vantagens e duas desvantagens para cada exemplo.
 2. (35 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 exemplo de teste utilizando 1-NN.
 - (c) Utilizando a matriz de distância explique a classificação dos exemplo de teste utilizando 3-NN sem peso.
 - (d) Utilizando a matriz de distância explique a classificação dos exemplo 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. (10 pontos) Utilize o classificador pelo vizinho mais próximo (1-NN) com distância euclidiana. Avalie este classificador fazendo metade dos exemplos de cada classe da base Iris como conjunto de teste e o restante como conjunto de treinamento. Calcule o número acertos no conjunto de teste. *Dica: você pode utilizar o `sklearn` scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html.*
 4. (15 pontos) Utilize os classificadores 7-NN com e 7-NN sem peso e avalie os classificadores 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 `sklearn` archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition.*
 5. (10 pontos) Faça o mesmo da questão anterior para a base Wine `archive.ics.uci.edu/ml/datasets/Wine`.
 6. (10 pontos) Faça o mesmo da questão anterior removendo a última coluna da base Wine