UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
COLEGIADO DE CIÊNCIA DA COMPUTAÇÃO

Relatório Trabalho 1 – Aprendizagem de Máquina

Título: KNNJS
Acadêmicos: Gilberto Antunes & Henrique Tomé Data: 21/03/2019

---

1. DESCRIÇÃO DO QUE É O TRABALHO

O objetivo do trabalho consiste em implementar um classificador baseado em vizinhanças (KNN). Para isso, foi implementada a entrada de dados com base em arquivo em formato .csv, o qual é utilizado nos experimentos.

Para que o processo tenha base para análise, deverão ser executadas 10 repetições. Os valores a serem comparados deverão ser os valores médios das 10 execuções.

2. DESCRIÇÃO DO CONJUNTO DE DADOS

O conjunto de dados utilizado durante os experimentos consiste em uma base composta por 690 instâncias, as quais são divididas em apenas duas classes, 1 e 2. Dentre os 690 exemplares do conjunto, 383 pertencem à classe 1 e 307 à classe 2. Cada instância é composta por 14 atributos, todos do tipo float.

3. DESCRIÇÃO PASSO-A-PASSO DO EXPERIMENTO

O primeiro passo consiste no carregamento do conjunto de dados. Para tanto, implementou-se uma rota que carrega o arquivo desejado em formato .csv e o armazena em uma lista encadeada de objetos, em que cada objeto corresponde a uma tupla contendo os atributos de cada instância. O último atibuto de cada objeto contém a classe daquele objeto.

O segundo passo consiste na divisão da base original em três subconjuntos mutuamente exclusivos: treino, teste e validação (Conforme apresentado na Figura 1). A instância que for designada para um conjunto não aparece nos outros.O conjunto de treino possui 50% do tamanho do arquivo original. Já as bases de validação e teste, tem 25% da dimensão, seguindo a estratégia de avaliação Hold-out.

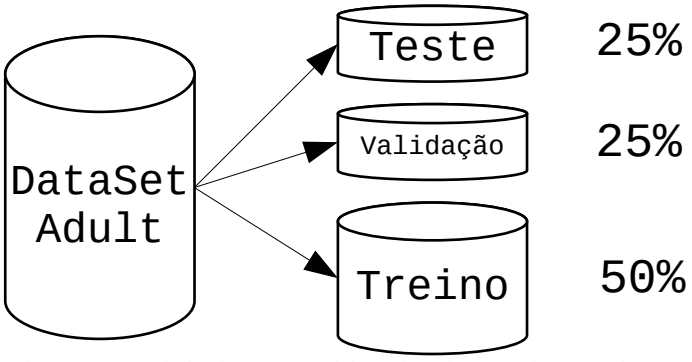

Figura 1: Divisão estratificada do conjunto de entrada.

No momento de separar a base original nos três conjuntos (treino, teste e validação), manteve-se as proporções originais das classes. Por exemplo, se um conjunto possui 200 instâncias da classe 1 e 100 da classe 2, o conjunto de treino terá 100 instâncias da classe 1 e 50 da classe 2.

As definições dos votos do classificador se deu por escolha do Tomé, o voto majoritário e voto ponderado por meio do método Inverso da distância Euclidiana.

A escolha das instâncias que formarão cada um dos conjuntos é totalmente aleatória. Assim, em cada execução do experimento, os conjuntos formados são diferentes. Depois de formados os conjuntos, o passo seguinte foi o treinamento do modelo de classificação. Como o KNN possui treinamento preguiçoso, esta etapa consiste apenas em armazenar o conjunto de teste. Para se determinar quais os melhores parâmetros dos métodos de classificação, adotou-se o conjunto de validação.

4. AVALIAÇÃO DOS EXPERIMENTOS

Depois de implementados os métodos de entrada, treinamento e escolha do melhor k, este sendo igual a 9 (k=9), foram realizados experimentos para avaliar alguns fatores dentro do classificador. Nestes experimentos foi comparado o desempenho do método ao se usar o Voto Majoritário Simples e o Voto Ponderado. Para o voto ponderado, foi implementada a estratégia do inverso da distância euclidiana.

Os resultados obtidos após as dez execuções são apresentados na Tabela 1 a seguir. Cada coluna apresenta a acurácia de uma das abordagens ao longo das 10 execuções. A última linha contém a Acurácia Média da estratégia e o seu devido Desvio Padrão.

Repetição	Voto Majoritário	Voto Ponderado
1	83,8150	83,2370
2	83,8150	83,8150
3	87,8613	87,2832
4	85,5491	85,5491
5	84,9711	84,9711
6	83,8150	81,5029
7	82,6590	82,0809
8	86,1272	86,1272
9	83,8150	83,2370
10	84,3931	83,8150
Média	**84,6821**	**84,1618**
DP	**4,4961**	**5,4348**

Tabela 1

5. ANÁLISE RESULTADOS

Observando-se os valores apresentados na Tabela 1 nota-se que odesempenho das abordagens implementadas é bastante similar. Todavia, é possível perceber que a abordagem do voto majoritário obteve a maior taxa de acerto, porém vale ressaltar que este resultado é situacional e em teste anteriores, houve casos em que o voto ponderado se sobressaiu em relação ao voto majoritário.