

Titulo do Artigo

Nome do Aluno¹

Vitória , Espirito Santo

^aUniversidade Federal do Espirito Santo

Abstract

O trabalho consiste na construção e análise de classificadores para o processo de detecção de classes de vinhos utilizando o wine dataset fornecido pela biblioteca sklearn. Dentro da proposta utilizamos os classificadores: Dummy Zero R (ZR), Naive Bayes Gaussian (NBG), K Nearest Neighbor (KNN), Decision Tree (DT), além da construção de um classificador baseado no uso do método KMeans, que o denominamos de: K Means Centroids (KMC). Para avaliação e comparação foram computadas métricas de acurácia dos classificadores. Cada classificador foi submetido a um processos de validação cruzada aninhada para a seleção dos melhores hiperparametros, e depois a multiplas rodadas de execução. Os resultados obtidos demonstraram que os classificadores KNN, KMC e NBG obtiveram os melhores resultados (com acurácias médias de 0.96, 0.96 e 0.97 respectivamente).

1. Introdução

O trabalho desenvolvido procurou utilizar de técnicas de aprendizado de máquina para o processo de classificação da base de dados wine fornecida pela biblioteca scikit learn. Para isso, lançamos de mão técnicas de classificação já difundidas, sendo elas:

- Naive Bayes Gaussian (NBG)
- K Nearest Neighbor (KNN)
- Decision Tree (DT)

Além disso, como proposta de desenvolvimento foi realizado a criação de um classificador baseado no método KMeans, que denominamos de K Means Centroids (KMC) explorando uma nova técnica de classificação. Ao final de tudo, utilizamos o classificador ingênuo Zero R (ZR) para realizar medidas comparativas.

Dessa forma, o artigo está organizado segundo a seguinte estrutura: 2 Base de Dados, explicando sobre o wine data set e suas características, 3 O Método KMC, com a descrição do método construído, 4 Descrição dos Experimentos Realizados e seus Resultados, demonstra os resultados obtidos pelos classificadores fazendo comparações de desempenho, 5 Conclusão, apresenta as conclusões obtidas com o trabalho e indicando futuras propostas.

2. Base de Dados

A base de dados wine consiste na análise química de vinhos produzidos na mesma região da Itália, contendo 178 instâncias de vinhos divididas em 3 classes que representam distintos cultivos, em que cada uma das instâncias contém 13 parâmetros.

2.1. Descrição do Domínio

Os dados extraídos são todos contínuos e com valores positivos.

2.2. Definição das Classes e das Características

O dataset é composto por 3 classes de diferentes cultivos, sendo extraídas por cada instância as seguintes características: 1) Álcool 2) Ácido málico 3) Cinzas 4) Alcalinidade das cinzas 5) Magnésio 6) Fenóis totais 7) Flavonóides 8) Fenóis não flavonóides 9) Proantocianinas 10) Intensidade da cor 11) Matiz 12) OD280/OD315 de vinhos diluídos 13) Prolina.

2.3. Número de Instâncias

O data set é composto por 178 instâncias divididas conforme a seguinte tabela:

| Classes | class 0 | class 1 | class 2 |
|------------|---------|---------|---------|
| Quantidade | 59 | 71 | 48 |

Table 1: Divisão das instâncias por classe do wine dataset

3. O Método KMC

Como parte da proposta do trabalho

4. Descrição dos Experimentos Realizados e seus Resultados

5. Conclusão

5.1. Análise geral dos resultados

5.2. Contribuições do Trabalho

5.3. Melhorias e trabalhos futuros

Referencias Bibliográficas

References