

Projeto weka 23/07/2021

Data de entrega: **06/08/2021**

Objetivo: Estudo das tarefas de agrupamento e classificação de dados empregando o software WEKA – modelo K-médias e rede neural MLP.

Etapas:

1. Descrição do software WEKA (disponível em https://[www.cs.waikato.ac.nz/ml/weka/).](http://www.cs.waikato.ac.nz/ml/weka/))
2. Descrição dos conjuntos de dados:
   1. “Iris Plant database” (Disponível em <http://www.hakank.org/weka/iris.arff>)
   2. “Wine recognition data” (Disponível em <http://www.hakank.org/weka/wine.arff>)
3. Descrição do modelo K-médias implementado no WEKA através da função de agrupamento “SimpleKMeans”.
   1. Detalhar a função dos parâmetros disponíveis para configuração do modelo.
   2. Identificar os parâmetros referentes ao número de grupos, número máximo de iterações e função de distância.
4. Executar os testes abaixo para os dois conjuntos de dados mencionados no item 2.
   1. Experimento 1: identificar o número de classes existentes no conjunto de dados e empregar o mesmo valor como número de grupos do modelo k-médias.
   2. Experimento 2: modificação do número de grupos – valores: {2, 3, 4, e 5}.
   3. Montar novos experimentos variando a semente (parâmetro “seed”) com o objetivo de avaliar a variação no resultado obtido a partir de centros iniciais diferentes.
5. Descrição do modelo MLP (Multi Layer Perceptron) implementado no WEKA através da função de classificação “MultilayerPerceptron”.
   1. Detalhar a função dos parâmetros disponíveis para configuração do modelo.
   2. Identificar os parâmetros referentes a taxa de aprendizado, momento e número de épocas.
6. Executar os testes abaixo para os dois conjuntos de dados mencionados no item 2. Para estas execuções deve ser selecionada a opção de teste “Cross-validation”, com 10 pastas conforme padrão do WEKA. Esta opção implementa a validação cruzada.
   1. Experimento 1: modificação da taxa de aprendizado – valores: {0,3; 0,6; 0,9}
   2. Experimento 2: modificação do momento – valores: {0,2 0,5; 0,8}
   3. Experimento 3: modificação do número de épocas: {500, 1000, 2000}
   4. Montar novos experimentos combinando taxa de aprendizado e momento. Os valores padrão do WEKA são 0,3 e 0,2, respectivamente. Defina alguns experimentos combinando os valores propostos nos itens a) e b). Não é necessário fazer todas as combinações.
7. Avaliar os resultados obtidos na etapa 4)
8. Avaliar os resultados obtidos na etapa 6), comentando cada resultado com relação à taxa de acerto do classificador. Para tanto, deve ser considerado o número de instâncias corretamente classificadas na saída produzida pelo WEKA.
9. Faça alguns testes modificando o número de pastas da validação cruzada. Empregue um número menor de pastas, entre 2 e 5, por exemplo, mantendo os valores dos outros parâmetros da rede neural no padrão do WEKA.
10. Gerar um relatório detalhando todas as etapas desenvolvidas.

Observações:

* Esta tarefa pode ser realizada em dupla e vale 4 pontos na nota do bimestre.
* Para as atividades realizadas em dupla, apenas um aluno deve enviar a solução no SIGAA (Não esqueça de inserir nomes e números de matrícula).
* Dúvidas: envie um email para [isadrummond@unifei.edu.br](mailto:isadrummond@unifei.edu.br)

