

## INF1771 - INTELIGÊNCIA ARTIFICIAL

### TRABALHO 3 – APRENDIZADO DE MÁQUINA

O objetivo do Trabalho 3 é desenvolver dois métodos de aprendizado supervisionado diferentes para realizar a tarefa de classificação em uma base de dados. Deve-se fazer ainda uma comparação entre os métodos implementados.

A base de dados pode ser escolhida no repositório da UCI (University of California, Irvine): <http://archive.ics.uci.edu/ml/>

No site da UCI, encontram-se algumas informações resumidas sobre as bases, mas, nos arquivos existem (quase sempre) informações completas a respeito das bases. Será responsabilidade de cada aluno escolher uma base com informações coerentes e adequadas para a aplicação dos métodos.

Para o desenvolvimento do sistema você deve seguir os seguintes passos:

- 1) Definir quais serão os atributos que serão usados para descrever os exemplos de treinamento. A classificação pode piorar ou melhorar dependendo do conjunto de atributos utilizado.
- 2) Se necessário, criar um programa para extrair os atributos e gerar um conjunto de treinamento e um conjunto de validação. Normalmente gera-se um único conjunto de dados e depois se divide ele em dois conjuntos (treinamento e validação). A proporção de dados utilizado para treinamento e validação é um aspecto a ser analisado no trabalho.
- 3) Utilizar 2 algoritmos de aprendizado supervisionado diferentes, treinando-os com o conjunto de treinamento e depois realizando a classificação do conjunto de testes para verificar qual algoritmo apresenta a melhor taxa de reconhecimento. Podem ser utilizados dois dos seguintes algoritmos:
  - Árvores de Decisão;
  - K-Nearest Neighbor (KNN) ;
  - Support Vector Machine (SVM) ;
  - Redes Neurais;
  - Outro de aprendizado supervisionado...

#### **Informações Adicionais:**

Durante a etapa de implementação dos algoritmos não é necessário implementar todos os classificadores para realizar os experimentos. É permitida a utilização de ferramentas de aprendizado de máquina, como por exemplo, o Weka (<http://www.cs.waikato.ac.nz/ml/weka/>) para a implementação. A ferramenta Weka trabalha com bases no formato ARFF. Portanto, vocês deverão se preocupar com a conversão da base para esse formato. Muitas dessas bases do repositório da UCI estão disponíveis em outros repositórios já no formato ARFF.

A implementação dos algoritmos pode ser feita em qualquer linguagem (C/C++, C#, Java...).

#### **Forma de Avaliação:**

Será avaliado se todas as etapas do processo foram cumpridas corretamente e a qualidade do método implementado. A avaliação também será baseada na apresentação dos resultados durante a aula e do

Relatório, que deverá conter:

- Descrição da modelagem dos exemplos de treinamento:
  - Atributos selecionados para descrever os exemplos;
  - Justificativa para a escolha dos atributos;
- Descrição dos experimentos realizados:
  - Variações no conjunto treinamento e testes;
  - Variação nos parâmetros dos algoritmos;
- Comparação dos algoritmos analisados:
  - Descrição de como os algoritmos escolhidos foram implementados;
  - Taxa de reconhecimento;
  - Tempo gasto no processo de treinamento;
  - Tempo gasto no processo de classificação de um exemplo desconhecido;
- Resultados Finais:
  - Escolha dos classificadores e conjunto de atributos;
  - Resultados alcançados pelo algoritmo que foi implementado (tempo gasto no processo de treinamento, tempo gasto no processo de classificação de um exemplo desconhecido, taxa de reconhecimento).
  - Conclusões finais;

Gráficos são bem vindos e necessários nas comparações dos algoritmos!

É muito importante experimentar os parâmetros dos métodos e das bases de treinamento.

### **Data de Limite de Entrega:**

11/12/2018

### **Forma de Entrega:**

Os trabalhos devem ser enviados até as 23:59h da data limite para o email [rcapua@gmail.com](mailto:rcapua@gmail.com). Na aula do dia 12/12 cada aluno/grupo deverá apresentar em sala os resultados obtidos. Cada apresentação deverá durar no máximo 15 minutos. No dia 11 de dezembro todos os alunos/grupo deverão enviar para o meu email um PDF contendo os slides que utilizarão na apresentação, assim como o relatório e o código-fonte.

Tanto na apresentação quanto no relatório, vocês deverão descrever a(s) base(s) utilizada(s), informar os algoritmos e parâmetros utilizados e analisar os resultados obtidos.

A entrega desse trabalho é opcional, desde que o aluno tenha entregado o trabalho 2 de prolog. Caso contrário, esse trabalho deve obrigatoriamente ser feito. Os alunos que não tiverem feito o trabalho 2 e não enviarem o trabalho 3 até a data limite receberão zero, assim como os alunos que não comparecerem a apresentação do trabalho 3. Não envie o trabalho após o prazo de modo algum. Uma vez enviado o trabalho, será atribuída uma nota ao mesmo que contará na média dos trabalhos. O plágio de trabalhos, seja parcial ou integral, implica em zero, sendo que essa nota contará na nota final dos trabalhos.

O trabalho deve ser feito em grupos de no máximo 2 pessoas.