

Especificação do 2º Trabalho Prático:

Comparação de Técnicas de Aprendizado de Máquina

1. Descrição e objetivo do trabalho:

Este trabalho prático consiste na aplicação de algoritmos de aprendizado supervisionado à diferentes bases de dados com o objetivo de comparar seus resultados.

O objetivo deste trabalho é treinar (ou induzir) 5 classificadores (ou algoritmos de regressão) diferentes com, no mínimo, 4 bases de dados diferentes e comparar os resultados obtidos. As bases de dados utilizadas podem ser obtidas no repositório UCI (<http://archive.ics.uci.edu/ml/>). É importante procurar por bases de dados completas (sem valores ausentes) e que sejam para tarefas de classificação (ou seja, tenham uma variável classe definida) ou regressão.

Os algoritmos sugeridos para a indução (aprendizado) dos algoritmos são: Árvore de Decisão, Vizinhos mais Próximos (K-Nearest Neighbor – KNN), Naive-Bayes (ou outro classificador Bayesiano, como o TAN, ou até o K2), Redes Neurais e Máquina de Vetor de Suporte (Support Vector Machines – SVM). Os algoritmos devem ser treinados usando-se uma estratégia de amostragem, como validação cruzada (cross validation), para uma avaliação melhor.

O trabalho poderá ser desenvolvido por **equipes de até 2 alunos**.

2. Informações Adicionais:

- Não é necessário implementar todos os algoritmos de aprendizado, apenas 1 deve ser escolhido para a implementação. Os outros algoritmos podem ser executados a partir de bibliotecas externas, como por exemplo, o *Weka* (<http://www.cs.waikato.ac.nz/ml/weka/>).
- Se necessário, desenvolva um pequeno programa para formatar os dados da base para que eles possam ser acessados mais facilmente durante o processo de treinamento e obtenção de resultados.

3. Entrega do Trabalho:

O **prazo final para a entrega** do trabalho é **dia 29/06/2018** (penúltima semana de aula). O trabalho deverá ser entregue através do Moodle, na disciplina de Inteligência Artificial. o arquivo compactado

(“.zip” ou “.rar”) contendo um **artigo em formato científico** (contendo entre 6 e 8 páginas, conforme o template sugerido em: <http://ijcai-17.org/FormattingGuidelinesIJCAI-17.zip>) que relate o desenvolvimento do trabalho e códigos fontes (mais o executável “exe”).

O texto científico deve conter (sugestão):

- Título relacionado ao tema do trabalho e identificação dos autores (como sugerido no formato).
- Introdução: que descreva o problema de aprendizado de máquina, visões, etapas e diferentes abordagens.
- Fundamentação teórica: que forneça base teórica sobre aprendizado supervisionado, os algoritmos envolvidos nos experimentos e suas características específicas (quando são adequados, quando não são, vantagens e desvantagens);
- Desenvolvimento do trabalho e análise de resultados: explique sobre as bases utilizadas e apresente os resultados obtidos pelos algoritmos para cada base de dados, comparando os algoritmos e destacando as características individuais de cada um que podem ter influenciado nos resultados.
- Finalmente, as conclusões e referências;

4. Forma de Avaliação:

O trabalho tem o **valor de 15% da nota total**. Será avaliada a implementação do algoritmo escolhido e a qualidade do artigo científico, observando-se a escrita correta, o conteúdo apresentado e a formatação adequada.

Obs: Se houver a detecção de plágio, o trabalho terá nota zero.