



ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
Departamento de Engenharia de Computação e Sistemas Digitais

PCS - 3838 Inteligência Artificial

3º Quadrimestre de 2020

2o. Exercício Prático

Utilize os algoritmos de AM baseados em (i) Árvore de Decisão e (ii) Redes Neurais para classificar e analisar/comparar os resultados por eles obtidos, utilizando um dataset público.

Os classificadores estão disponíveis no repositório AIspace, produzido no Laboratory for Computational Intelligence at the University of British Columbia, Canadá. Vocês devem utilizar o executável em Java, por permitir uma melhor visualização. Os arquivos estão disponíveis em:

- Decision Tree: <http://www.aospace.org/dTree/version4.4.0signed/dTree.jar>
- Neural Networks: <http://www.aospace.org/neural/version4.3.8signed/neural.jar>

Os datasets devem ser obtidos do Machine Learning Repository, mantido pelo Center for Machine Learning and Intelligent Systems at University of California, Irvine, disponíveis em <https://archive.ics.uci.edu/ml/datasets.php>. Escolham datasets com as seguintes propriedades:

- Tarefa: classification
- Tipo de dado: multivariate
- Número de Atributos: menos de 10 (vocês devem escolher datasets cujo número de atributos variem entre 4 a 6)

Em ambos os casos, separe convenientemente os dados entre aqueles usados para treinamento (70%) e os usados para teste (30%). Calcule as seguintes métricas:

- Precisão: $TP/(TP+FP)$
 - qual a porcentagem de itens classificados na classe C que realmente pertencem a esta classe?
- Recall: $TP/(TP+FN)$
 - qual a porcentagem de itens da classe C que foram classificados como tal?

Utilize estes valores para decidir quando deve terminar o treinamento.

Algumas dicas:

- Na geração da árvore de decisão, utilize o ganho de informação para a escolha do melhor atributo. Escolha um valor mínimo deste ganho como critério de parada, baseado no seu dataset. Executar o algoritmo, utilize a opção show plot para verificar os erros de classificação (utilize o erro quadrático). Altere o valor caso as métricas de precisão e recall não sejam adequadas.
- Na geração da rede neural, inicialize os pesos de forma randômica. Ajuste a taxa de aprendizado para um valor conveniente (entre 0.1 e 0.2). Utilize apenas uma camada escondida, adote um número inicial de neurônios nesta camada (lembres-e que limitar no máximo em $2n+1$, sendo n o número de neurônios na camada de entrada). Altere o valor caso as métricas de precisão e recall não sejam adequadas.



ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
Departamento de Engenharia de Computação e Sistemas Digitais

Em ambos os classificadores, salve o seu dataset (Save Problem), que irá gerar um arquivo xml. Também fotos da execução que julgue relevantes para o seu relatório, contendo a árvore gerada ou a rede treinada, taxas de erro etc.

O exercício deve ser realizado ***em duplas***: imagina-se que um aluno irá implementar/testar a árvore de decisão, o outro irá implementar/testar a rede neural e ambos farão a análise comparativa conjuntamente. Cada dupla deverá escolher um dataset **distinto**, a ser validado pelo professor.

A avaliação levará em conta (i) a entrega dos dois arquivos xml e sua corretude e (ii) um relatório descrevendo o trabalho e análise comparativa dos resultados.

Sugerem-se as seguintes seções para o relatório:

1. *Introdução*, contendo uma breve descrição da tarefa e ferramentas/repositórios utilizados;
2. *Descrição do dataset*, contendo uma breve descrição dos dados nele contidos;
3. *Resultados obtidos*, apresentando os resultados de classificação das duas abordagens;
4. *Análise dos resultados*, comparando os resultados obtidos pelas abordagens;
5. *Conclusões e trabalhos futuros*, identificando o resultado do trabalho e suas possíveis extensões.

A entrega deve ser realizada até o dia **08/12/20, às 17:00**, através do Moodle da disciplina. Ambos os alunos da dupla devem entregar o exercício de forma idêntica. Devem ser entregues dois arquivos:

- a) Arquivo <3838_ex2_20_nusp_primeironome>.zip, contendo os dois arquivos xml gerados para cada classificador, cujos nomes deverão ser: <nome-do-dataset>_dt.xml e <nome-do-dataset>_nn.xml;
- b) Arquivo <3838_ex2_20_nusp_primeironome>.pdf, contendo o relatório.