

Enunciado do Trabalho 2

O web site oficial da ferramenta Netica da Norsys disponibiliza diversas redes bayesianas já definidas para download (<http://www.norsys.com/netlibrary/index.htm>), em diversas categorias do conhecimento especialista: *Environmental, Medical, Financial/Management, Geological, Industrial Diagnosis, Other Industrial*. Para cada uma dessas categorias, existem várias redes-exemplo e descrição da fonte explicativa do problema.

A tarefa é escolher uma dessas redes e proceder com um experimento em Aprendizado de Máquina, descrito a seguir:

- 1- Eleger uma característica de classe;
- 2- Utilizar técnica de amostragem em redes bayesianas para produzir um dataset de exemplos;
- 3- Reservar parte do dataset para treinamento e parte para testes;
- 4- Treinar modelo Naïve Bayes;
- 5- Treinar modelo de Árvores de Decisão (ou Floresta Aleatória);
- 6- Testar modelos com o conjunto de testes;
- 7- Responder a algumas perguntas: O quanto cada modelo se aproxima da rede bayesiana original? O quanto o tamanho do dataset gerado influencia na qualidade? Smoothing se faz necessário? Quem possui melhor comportamento: Naïve Bayes ou Árvore de Decisão?
- 8- Concluir o experimento.

OBSERVAÇÕES:

- 1- É imperativo o reuso de códigos da disciplina, disponibilizados particularmente nas pastas compartilhadas 12racprobiA, 15NaiveBayes e 16arvoresdecisaolA. O código DS&W... dentro da pasta 12racprobiA/Eliminacao..., desenvolvido por alunos da turma passada, é, por exemplo, uma ótima referência e pode ser reutilizado também; o código traz inclusive um parser para fazer a leitura dos arquivos no formato DSC do Netica;
- 2- Existem algumas técnicas estudadas na literatura relacionada para realizar amostragem: *Prior Sampling, Rejection Sampling, Likelihood Weighting, Gibbs Sampling* são algumas delas (ver seção 14.5 do livro). Pode-se buscar por um código disponível na Internet e incorporá-lo/adaptá-lo nas codificações citadas anteriormente;
- 3- O relatório final de experimentação deve ser entregue em um único arquivo no formato .ipynb contendo naturalmente toda a codificação; faça a compactação desse arquivo com os datasets criados em um arquivo formato .zip e envie por email (utilize o nome da rede Netica escolhida como nome do arquivo);
- 4- Uso de gráficos para facilitar comparação de resultados serão valorizados;
- 5- A equipe deve ser composta por 6 (seis) integrantes para que o trabalho seja de fato bem dividido e dimensionado; a orientação é de que uma dupla trate das amostragens e geração de dados, a segunda dupla trate dos modelos Naïve Bayes, e a terceira dupla trate dos modelos de Árvores de Decisão. Um integrante deve cuidar das reuniões de acompanhamento com o professor.

Integrante 1 (Amostragem e dados): _____

Integrante 2 (Amostragem e dados): _____

Integrante 3 (Naïve Bayes): _____

Integrante 4 (Naïve Bayes): _____

Integrante 5 (Árvores de Decisão): _____

Integrante 6 (Árvores de Decisão): _____