

CPE 722 - Instruções sobre o trabalho.

O trabalho é individual.

1 - Escolha o banco de dados. Pode ser um banco que você já tenha ou selecione um de

<http://archive.ics.uci.edu/ml/datasets.html>

default task: classification attribute type: numerical or categorical

A dimensão da entrada deve ser igual ou maior que 5 e sugerimos que seja menor que 15. (o ideal é entre 5 e 10 entradas). Escolha um tema que lhe interesse, mas não escolha um com dimensão de entrada muito grande, porque neste caso você terá muito trabalho.

Não serão aceitos temas repetidos. Após escolher seu tema me comunique por e-mail. Se alguém já houver escolhido o tema anteriormente eu avisarei e você deverá escolher outro.

Manterei na página uma tabela com os nomes e temas escolhidos. Consulte a tabela antes de me enviar seu tema, para verificar que ninguém o escolheu.

2 - Se o número de instâncias for muito elevado escolha um subconjunto com cerca de 200 a 500 instâncias, tomando cuidado que todas as classes esteja representadas nele.

3 - Use ao menos duas técnicas de clusterização com o mesmo banco de dados, uma supervisionada e outra não supervisionada. Processe cada técnica de forma completamente independente.

4 - Escolha das variáveis a usar. Os processamentos para aprendizado supervisionado (correlação) e não supervisionado (histograma de distâncias) são diferentes. Use ambos.

5 - Faça a normalização das variáveis de modo a obter classes aproximadamente esféricas. Os métodos para normalização para treinamento supervisionado e não supervisionado são diferentes. Use ambos.

6 - Monte o conjunto de treinamento e também um conjunto para o teste. Não utilize entradas com dados faltantes, a menos que você esteja estudando seu efeito..

7 - Apresente as tabelas de confusão para o conjunto de treinamento e o de teste e comente. Compare os resultados obtidos pelos dois métodos.

8 - Você pode incluir mais algum estudo adicional que julgar interessante (efeito de dados faltantes, relevância de entradas, etc.). Isto será valorizado.

Apresentação:

O trabalho deve ser apresentado apenas em versão eletrônica. O texto pode ser composto pelos slides que serão apresentados, desde que contenha todos os resultados das etapas descritas acima.. O texto deve ser completo mas conciso.

Prepare duas versões eletrônicas, uma em *.doc e outra em *.pdf. O nome do arquivo é o seu nome, jose_silva.doc, jose_silva.pdf. As versões eletrônicas devem ser enviadas para mim até 01/12/2014, independente do dia da apresentação. Atrasos serão considerados no conceito.

O trabalho deverá ser apresentado em 15 minutos, após o que haverá uma sessão de perguntas aberta a todos, como em um congresso.

A presença de todos é esperada em todas as apresentações, porque esta troca de experiências é importante e complementa o aprendizado. Estão dispensados para as apresentações de 8 às 10 h, evidentemente, aqueles que tem motivos aceitáveis (outra aula, trabalho, etc.).