

### Lab 3

**Objetivo:** O objetivo deste Lab é realizar treinamentos utilizando o RSNNS. Portanto, as questões devem ser respondidas a partir da utilização da biblioteca RSNNS no R. Logo, deve-se enviar, além das análises, também o código R para cada questão desenvolvida.

**Entrega:** Até as 23h59min do dia 10/05/2012.

**Formato da entrega:** Idêntico ao Lab anterior (se não sabe, releia). Ao escrever seu relatório, lembre-se de: (1) incluir artefatos (imagens, dados, etc.) que comprovem sua resposta, (2) tornar os passos utilizados fáceis de entender e replicar.

**Bases de Dados:** Neste exercício serão utilizadas duas bases de dados. A primeira [**crx.csv** e **crxPCA.csv**] refere-se a transações efetuadas com cartões de crédito. Apesar da base ser real, todos os dados foram modificados para preservar o sigilo dos usuários. A base contém dados do usuário, das transações e o referido resultado (se a transação deveria ter sido aprovada ou rejeitada). A base é mista, logo contém dados categóricos e numéricos. A segunda base [**moviebase.csv**] contém 100 mil votações de usuários sobre diversos filmes. Os filmes são rotulados quanto ao gênero e cada usuário aplicou um escore variando de 1 a 5 para cada filme que opinou. Na base fornecida, os dados foram reunidos em um único arquivo e foi adicionada uma coluna referente à classificação do filme, sendo "Great" se o filme tiver recebido nota maior ou igual a 3 e "Bad" caso contrário.

#### Ajuda em RSNNS:

Ao contrário do Weka, no qual boa parte do processo de treinamento é realizado automaticamente pela ferramenta, no RSNNS praticamente todo o processo precisa ser especificado por você. Portanto, é essencial conhecer os comandos básicos e, acima de tudo, compreender bem o processo de treinamento.

Aqui serão listados os comandos e procedimentos básicos para um treinamento simples em MLP usando o RSNNS.

1. *Instalação:* `install.packages('RSNNS')`.
2. *Carregamento da biblioteca:* `library('RSNNS')`
3. *Carregamento dos dados:* Basta carregar os dados em R por intermédio do comando `read.csv` ou similar. Cada linha corresponde a uma instância e cada

coluna a uma característica de cada instância. Atenção: O Weka converte automaticamente os dados categóricos em dados numéricos. No RSNNS, você precisa realizar esta operação manualmente (estude o comando `decodeClassLabels`).

4. *Divisão dos conjuntos*: Você também precisa dividir os conjuntos entre treinamento e teste (estude o comando `splitForTrainingAndTest`).

5. Treinamento: use alguma função de treinamento. Procure por `mlp`, `som`, `elman`, etc.

6. *Avaliação*: existem funções prontas para traçar a evolução do erro de treinamento (`plotIterativeError`), para mostrar os erros de regressão ou classificação (`plotRegressionError`), para traçar a curva ROC (`plotROC`), para determinar a matriz de confusão (`confusionMatrix`) dentre outras.

Exemplos de treinamento podem ser encontrados em

<http://dicits.ugr.es/software/RSNNS/index.php?view=Examples%20of%20high-level%20API>.

Mais detalhes sobre todas as funções podem ser encontrados em

<http://cran.r-project.org/web/packages/RSNNS/RSNNS.pdf>.

**Atenção!** Para avaliar se um treinamento é melhor que outro, decida bem qual medida de desempenho irá utilizar!

1) Utilizando a base de dados **crx** [**crx.csv**], desenvolva um código para realizar um treinamento usando a função *MultilayerPerceptron* tendo como valores para o treinamento, todos os valores padrão do RSNNS. A base já foi preparada para o RSNNS utilizando-se a função *NominalToBinary* do Weka. (Tempo estimado para treinamento: < 1 min).

- Você deve entregar o script R utilizado.
- No seu documento de análise deve estar o raciocínio para a escolha realizada.

2) Pode-se afirmar, com 95% de confiança, que existem diferenças significativas na realização do treinamento com diferentes valores da taxa de aprendizagem (Learning Rate)? Teste as taxas 0.01, 0.11, 0.21, 0.31, 0.41, 0.51, 0.61, 0.71, 0.81 e 0.91. Não esqueça de repetir o experimento 10 vezes. Por padrão a inicialização dos pesos já é aleatória no RSNNS. (Tempo estimado para treinamento: 20 min).

- Você deve entregar o script R utilizado.
- No seu documento de análise deve estar suas conclusões.

3) Pode-se afirmar, com 95% de confiança, que o uso da técnica PCA (resultado do processamento PCA disponível no arquivo **crxPCA.csv**) pode produzir melhores resultados que o treinamento sem a utilização do PCA? Use a melhor taxa de aprendizagem encontrada na questão 2. (Tempo estimado para treinamento: 10 min).

- Você deve entregar o script R utilizado.
- No seu documento de análise deve estar a sua conclusão.

4) Usando a base de filmes [**moviebase.csv**], treine uma rede que produza uma função de regressão para estimar a nota de um filme com base nos escores de todos os usuários.

- Você deve entregar o script R utilizado.
- No seu documento de análise devem estar todos os resultados que comprovem se seu treinamento é adequado ao problema.

5) Se ao invés de uma regressão, fosse feita uma classificação, dadas duas classes: "Great" e "Bad" sendo ambas obtidas a partir das notas, o resultado de classificação seria melhor ou pior que o da regressão?

- Você deve entregar o script R utilizado.
- No seu documento de análise devem estar os resultados que comprovem sua escolha.

6) Utilizando a função **som** para a obtenção de mapas auto-organizáveis, escolha características da base de filmes [**moviebase.csv**] e tente identificar semelhanças entre os agrupamentos resultantes. Siga os passos:

- 1 - Gere uma rede 3x3 ou 5x5;
- 2 - Para cada instância, identifique o neurônio vencedor (ver `saveWinnersPerPattern`);
- 3 - identificar semelhanças para um mesmo neurônio vencedor. A análise pode ser qualquer análise estatística de sua preferência.  
(Tempo estimado para treinamento da rede 5x5: 10 min)

*Obs.: Recomenda-se utilizar o R 2.15.0 32-bit pois a versão 64-bit pode apresentar problemas quanto ao resultado.*

- Você deve entregar o script R utilizado.
- No seu documento de análise deve estar o raciocínio para a escolha realizada e também sua análise quanto aos agrupamentos encontrados.