



Curso de Engenharia de Sistemas e Informática - 5º Ano

Análise Inteligente de Dados

Ficha de Trabalho N.º 11

Objectivo: Estudo das Redes Neurais (MLP - Multi-Layer Perceptron) e das Árvores de Decisão utilizando o algoritmo Build C5.0. Análise comparativa das duas abordagens e utilização conjunta.

I – Redes Neurais.

Vamos prever o campo “Resposta à Campanha” utilizando uma rede neuronal.

1. Comece com um painel Stream limpo. Coloque um nó Source (ou do tipo Fich. Variável ou SPSS Import) no painel Stream e ligue-o ao ficheiro utilizado na ficha anterior charity.csv ou charity.sav, de acordo com o formato relevante para o nó fonte escolhido. No caso de escolher o nó Fich. Variável, leia os nomes dos campos do ficheiro. No caso do fich. SPSS, utilizar os valores e labels das variáveis.
2. Ligue um Nó Type e Table ao nó Fonte, criando um Stream. Execute o Stream, permitindo que o Clementine defina automaticamente os tipos dos campos.
3. Edite o nó Type. Coloque a direcção de todos os campos para NONE.
4. Tentaremos prever a resposta à campanha utilizando os campos abaixo indicados. Alterar a direcção de todos os cinco campos abaixo indicados para **IN** e o “resposta à campanha” para **OUT**.
Pre-campaign expenditure
Pre-campaign visits
Gender
Age
Mosaic Brands (que deve ser alterado para tipo Auto Set)
5. Adicione o **Nó Train Net** e ligue-o ao nó Type. Execute o Nó Train Net com os parâmetros sugeridos pelo Clementine.
6. Depois de terminar o treino do modelo, inspecione o Nó Net gerado na paleta Generated Models. Qual é a precisão predita da rede neuronal? Quais são os campos mais relevantes da rede neuronal?
7. Coloque o Nó Net gerado no painel Stream e ligue-o ao nó Type. Ligue-lhe em seguida um nó Table e execute essa parte do stream. Investigue os campos da tabela e interprete os novos campos adicionados.
8. Ligue-lhe seguidamente ao nó Net um Nó Matrix e crie a matriz de dados das resposta real à campanha face à resposta predita pela rede neuronal. Que grupo corresponde àquele que o modelo mais correctamente prediz?
9. Utilize alguns dos métodos gráficos estudados, tais como gráficos web e histogramas, de forma a que consiga entender o raciocínio que está por trás das predições da rede neuronal. Descreva as conclusões chegadas.
10. Inspecione o ganho e rácio do ganho (lift) do modelo gerado utilizando um nó Evaluation Chart. Interprete os resultados obtidos.
(Obs. A vantagem de ganho avalia-se em termos do afastamento da linha que representa o ganho do modelo face à linha base e quão mais rápido se afastar no início – mais próxima de um gráfico tipo step. Quanto ao gráfico lift (razão da percentagem

- de registos em cada quantil acertados, divididos pela percentagem total de acertos nos dados de treino – vantagem avaliada como uma razão da taxa base).
11. Guarde uma cópia do trabalho desenvolvido sob o nome Rede_Neuronal.str

II - Árvores de Decisão

Vamos tentar prever o campo “Resposta à campanha” utilizando a técnica de indução de regras C5.0, no fundo, o mesmo propósito traçado na secção anterior.

1. Carregue o stream gerado na secção anterior, se não estiver carregado ainda.
2. Ligue um **Nó Build C5.0** ao nó Type e execute o stream, utilizando os parâmetros sugeridos pelo Clementine.
3. Depois do modelo gerado, inspecione o modelo criado na paleta Modelo Gerados (**C5.0 Rule**). Expanda completamente as regras geradas de forma a compreender o processo de decisão indicado pelo modelo. Da análise feita parece-lhe que a árvores de decisão se comporta de forma semelhante à rede neuronal criada na secção anterior? Fundamente a sua resposta.
4. Nas regras geradas interprete o significado dos número à frente de cada regra (se não aparecerem, actue por forma a que surjam).
5. Efectue uma comparação entre o valor real da resposta à campanha face o predito e respectiva confiança.
6. Faça gerar e verifique um conjunto de regras e interprete-as.
7. Ligue um **Nó Matrix** ao **Nó C5.0 Rule** e crie uma matriz de dados da resposta real à campanha face à resposta predita pela Árvore de Decisão. Parece-lhe que este modelo tem uma precisão superior à rede neuronal do ponto anterior?
8. Faça uma avaliação mais fundamentada através do nó Evaluation Chart e compare o resultado com o correspondente relativo ao modelo de rede neuronal.
9. Guarde uma cópia actualizada do stream.

III - Combinar Redes Neurais e Árvores de Decisão.

Já que os propósitos genéricos dos dois métodos estudados nesta ficha são similares, vamos comparar os dois modelos criados nas duas secções anteriores e depois utilizar também a combinação para tentar compreender melhor a forma como a rede neuronal faz as suas predições.

1. Carregue o stream actualizado na secção anterior, se não estiver carregado ainda.
2. Efectue arranjos ao stream de forma a que o modelo de rede neuronal e o da árvore de decisão fiquem no mesmo stream, ligados ao nó Type.
3. Coloque um nó **Analysis** no final desse stream e execute-o. Qual dos modelos tem o maior grau de precisão? Até onde são consistentes nas suas decisões?
4. Inspeccione o modelo da Árvore de Decisão gerado e crie de forma automática um nó Filter.- Ligue esse nó criado ao nó Type e edite-o. O modelo da árvore de decisão utiliza todos os campos de entrada disponíveis? Indique possível utilização em termos de utilização combinada com outro modelo.
5. Ligue um nó Type ao nó **Net** que já está no stream e retire o nó C5.0 Rule do stream.
6. Edite o nó **Type** e coloque a NONE a direcção dos campos “response to campaign” e “\$NC-Response” (para que não sejam utilizados pelo algoritmo que iremos utilizar para descrever a forma como a predição é feita na rede neuronal criada).
7. Coloque o campo “\$N-Response to campaign” em OUT (o campo de saída – de que

queremos conhecer as caracterizações casuísticas).

8. Ligue um nó **Build C5.0** ao nó Type anterior e **execute** o stream. Teça algumas considerações acerca da compreensão do modo como a rede neuronal faz as previsões?