

REDES NEURAIIS - Exercício com o WEKA

Problema de Classificação de Padrões

1º Trabalho

Uma instituição financeira possui uma base de dados com o histórico de crediário oferecido aos seus clientes. Baseado neste histórico, a instituição deseja inferir se um novo cliente pagará ou não a dívida contraída.

A base de dados possui 2077 exemplos, com 11 atributos cada, de créditos concedidos aos seus clientes. A base informa ainda se o cliente honrou ou não o pagamento do empréstimo. Abaixo segue a descrição dos atributos da base:

	Nome das Variáveis	Descrição	Tipo	Valores possíveis
1	ESTC	E stado civil	Categórica	0,1,2,3
2	NDEP	Número de d ependentes	Categórica	0,1,2,3,4,5,6,7
3	RENDA	R enda Familiar	Numérica	300-9675
4	TIPOR	T ipo de residência	Categórica	0,1
5	VBEM	Valor do b em a ser adquirido	Numérica	300-6000
6	NPARC	Número de p arcelas	Numérica	1-24
7	VPARC	Valor da p arcela	Numérica	50-719
8	TEL	Se o cliente possui t elefone	Categórica	0,1
9	IDADE	I dade do cliente	Numérica	18-70
10	RESMS	Tempo de moradia (R esidência) (em m eses)	Numérica	0-420
11	ENTRADA	Valor da e ntrada	Numérica	0-1300
=	CLASSE	=1 se o cliente pagou a dívida	Categórica	0,1

A partir da base original, foram criadas 3 bases de treinamento, com 1500 exemplos cada escolhidos aleatoriamente a partir da base original, e 3 bases de testes com 577 exemplos cada, representando, respectivamente, 72,2% e 27,8% do total de cada sub-grupos de dados. Estas bases estão nos arquivos *treino01.txt*, *treino02.txt*, *treino03.txt*, *teste01.txt*, *teste02.txt* e *teste03.txt*.

Utilizando o software WEKA, crie um classificador, baseado em redes neurais, capaz de informar se um novo cliente será potencialmente adimplente ou não. Para isso, siga as instruções abaixo:

- I) Abra cada um dos arquivos no WEKA e grave-os em formato *.arff*. Através de um editor de textos (por exemplo, o WordPad), altere o tipo associado às variáveis categóricas conforme o exemplo abaixo:

@attribute NDEP **numeric** ⇒ @attribute NDEP {0,1,2,3,4,5,6,7}

Para cada uma das configurações abaixo, apresente os resultados para cada par de conjuntos de treino e de teste, assim como a média e o desvio padrão dos 3 pares:

- II. **Sem normalização** dos atributos de entrada.
- III. **Com normalização** dos atributos de entrada e com 2 números diferentes de neurônios na camada escondida. Explique a escolha dos números.
- IV. **Com normalização** dos atributos de entrada e variando o número de épocas durante a fase de treinamento. Escolha 3 durações de treino diferentes (por exemplo: 1, 100 e 1000).
- V. **Com normalização** dos atributos de entrada e utilizando um **conjunto de validação**.
- VI. **Com normalização** dos atributos de entrada e **codificando binariamente** os atributos de entrada.
- VII. Tente obter melhores resultados (se possível) agrupando algumas categorias das variáveis **ESTC** e **NDEP**. Para isto, utilize o filtro não supervisionado **MergeTwoValues**.

Para os itens II, IV, V, VI e VII, indique para cada um dos casos o número de neurônio na camada escondida e explique a sua escolha. Para todos os itens, não varie a taxa de aprendizagem nem o termo de momento.

O relatório deve apresentar o resumo e uma análise dos resultados obtidos. Qualquer dúvida, entrar em contato com o monitor: Leonardo Forero (mendoza@ele.puc-rio.br)

Sugestão: Caso queira poderá utilizar a tabela para visualizar os resultados.

[illegible]