Pontifícia Universidade Católica do Rio de Janeiro

Departamento de Engenharia Elétrica

Redes Neurais – Prof<sup>a</sup>. Marley Vellasco

## 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. A Tabela 1 descreve os atributos da base.

Tabela 1 – Descrição dos atributos do problema.

	Nome das	Descrição	Tipo	Valores possíveis				
	Variáveis							
1	ESTC	Estado Civil	Categórica	0, 1, 2, 3				
2	NDEP	Número de <b>dep</b> endentes	Categórica	0, 1, 2, 3, 4, 5, 6,				
				7				
3	RENDA	<b>Renda</b> Familiar	Numérica	300 - 9675				
4	TIPOR	Tipo de Residência	Categórica	0, 1				
5	VBEM	Valor do <b>bem</b> a ser adquirido	Numérica	300 - 6000				
6	NPARC	Número de parcelas	Numérica	1 - 24				
7	VPARC	Valor da <b>parc</b> ela	Numérica	50 – 719				
8	TEL	Se o cliente possui <b>tel</b> efone	Categórica	0, 1				
9	IDADE	Idade do cliente	Numérica	18 - 70				
10	RESMS	Tempo de moradia	Numérica	0 - 420				
		( <b>Res</b> idência, em <b>m</b> ese <b>s</b> )						
11	ENTRADA	Valor da <b>entrada</b>	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. 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 pagará ou não a dívida. Para isso, siga as instruções abaixo.

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:

- a) Sem normalização dos atributos de entrada.
- b) Com normalização dos atributos de entrada e com 2 números diferentes de neurônios na camada escondida. Explique a escolha dos números.
- c) 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).
- d) Com normalização dos atributos de entrada e utilizando um conjunto de validação.
- e) **Com normalização** dos atributos de entrada e **codificando binariamente** os atributos de entrada.
- f) 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 **a**, **c**, **d**, **e**, **f**, escolha para cada um dos casos o número de neurônios 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.

Sugestão: Caso queira, poderá utilizar as tabelas abaixo para visualização dos resultados.

Tabela 2 – Resultados obtidos para cada configuração

	Grupo I						Grupo II					Grupo III						
	Configuração						Configuração					Configuração						
	I	II	III	IV	V	VI	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Hidden Layer																		
Hidden Layer																		
Class. Correta em %																		
Class. Incorreta em %																		
MAE																		
RMSE																		
RAE (%)																		
RRSE (%)																		

Tabela 3 – Média e Desvio padrão de cada configuração

	Média							Desvio Padrão						
	Conjunto							Conjunto						
	I	II	III	IV	V	VI	I	II	III	IV	V	VI		
Hidden Layer														
Hidden Layer														
Class. Correta em %														
Class. Incorreta em %														
MAE														
RMSE														
RAE (%)														
RRSE (%)														