**Estudo dirigido RNAs e classificadores**

Felipe Stadler, RA: 17021426

**1) Considere um perceptron simples com 5 entradas (4 elementos de dados e um bias = 1). Suponha que o conjunto de pesos atuais é (-0.5, 0.7, 0.6, 0.1, -0.2) e que os seguintes exemplos devem ser processados:**



**Qual os valores dos pesos após o processamento destas entradas? (Assuma que a taxa de aprendizagem é de 0.5).**

(-0,5) \* 1 + 0,7 \* 1 + 0,6 \* 1 + 0,1 \* 1 + (-0,2) \* (-1) = 1,1

Saída = 1 e <W X> > 0, não altera os pesos.

(-0,5) \* 1 + 0,7 \* (-1) + 0,6 \* (-1) + 0,1 \* 1 + (-0,2) \* 1 = -1,9

Saída = 1 e <W X> <0, alteramos os pesos:

W0 = W0 + X0 \* taxa\_aprend = -0,5 + 1 \* 0,5 = 0

W1 = W1 + X1 \* taxa\_aprend = 0,7 + (-1) \* 0,5 = 0,2

W2 = W2 + X2 \* taxa\_aprend = 0,6 + (-1) \* 0,5 = 0,1

W3 = W3 + X3 \* taxa\_aprend = 0,1 + 1 \* 0,5 = 0,6

W4 = W4 + X4 \* taxa\_aprend = -0,2 + 1 \* 0,5 = 0,3

Após o processamento destas entradas, os pesos são: (0, 0.2, 0.1, 0.6, 0.3)

**2) Considere archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(28diagnostic), utilize o software WEKA para treinar um rede neural de múltiplas camadas para classificar os casos contidos na base de dados. Use redes com uma camada intermediária, uma camada de entrada e uma cada de saída. Durante os testes, utilize redes com as seguintes quantidades de nós na camada intermediária: 10, 15 e 20 . No treinamento de cada configuração rede realize testes com 200, 400 e 600 épocas.**

**Obs: - Utilize validação cruzada com 10 folds.**

**a) Para cada um dos testes determine qual das seguintes taxas de aprendizagem teve o melhor desempenho : (0.1, 0.3, 0.5)**

Camadas Intermediárias: 10, 15 e 20

Número de Épocas: 200 e 400 e 600

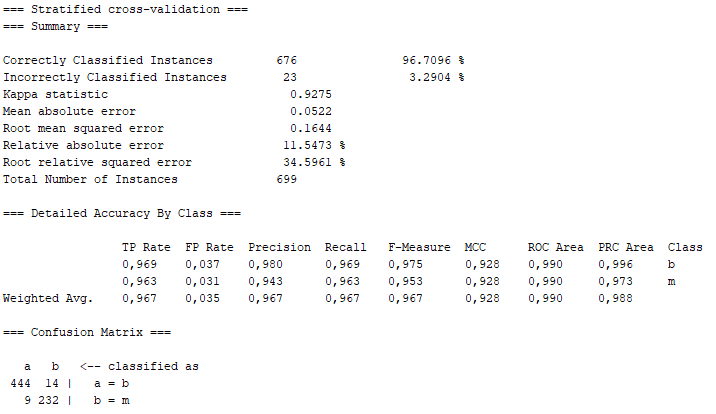
Taxas de aprendizagem: 0.1, 0.3 e 0.5

Resultados utilizando validação cruzada com 10 folds:

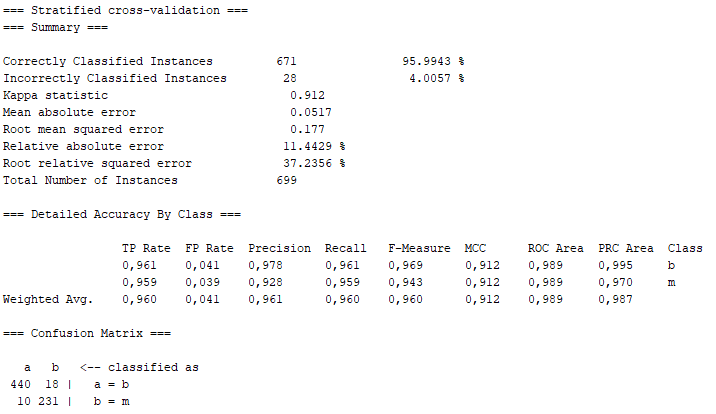
Os valores nos itens indicam, respectivamente:

- Camadas Intermediárias x Número de Épocas x Taxa de Aprendizagem:

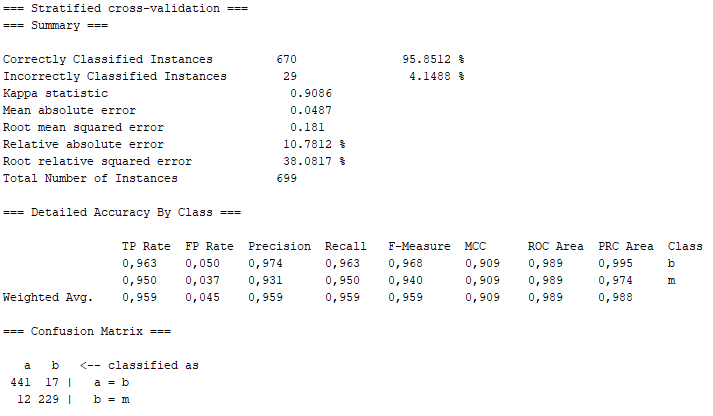
- 10 x 200 x 0.1:



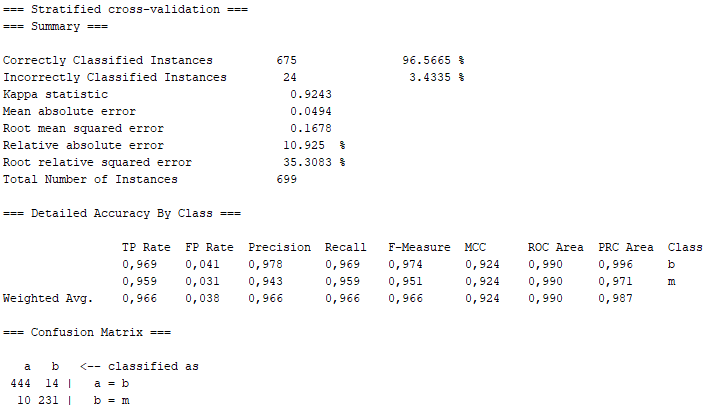
- 10 x 400 x 0.1:



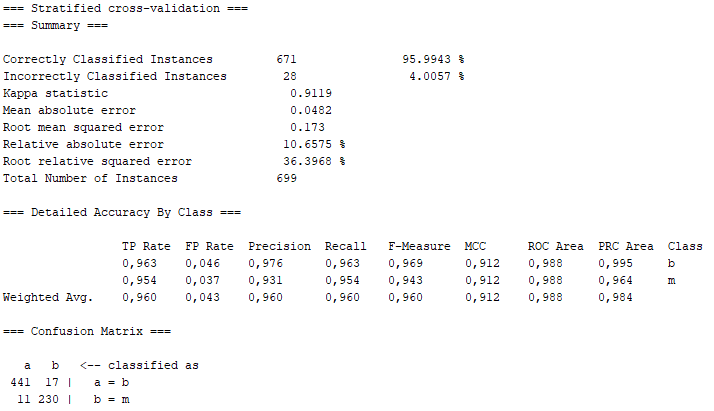
- 10 x 600 x 0.1:



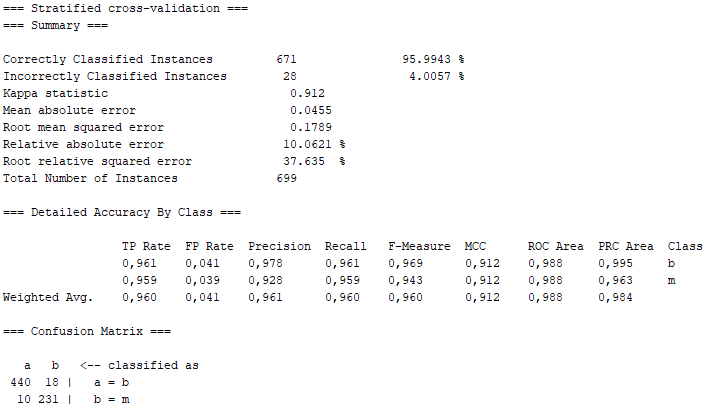
- 15 x 200 x 0.1:



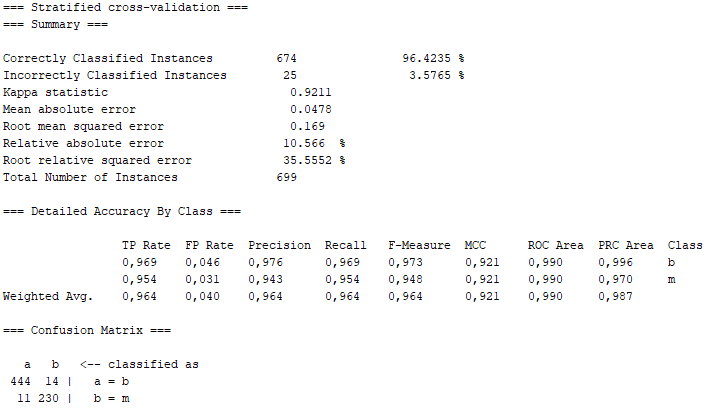
- 15 x 400 x 0.1:



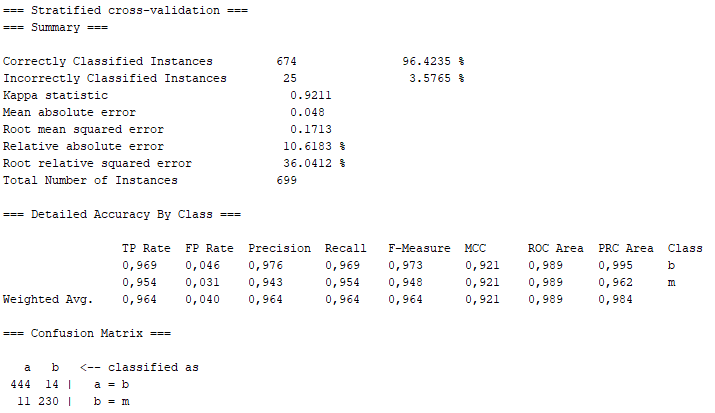
- 15 x 600 x 0.1:



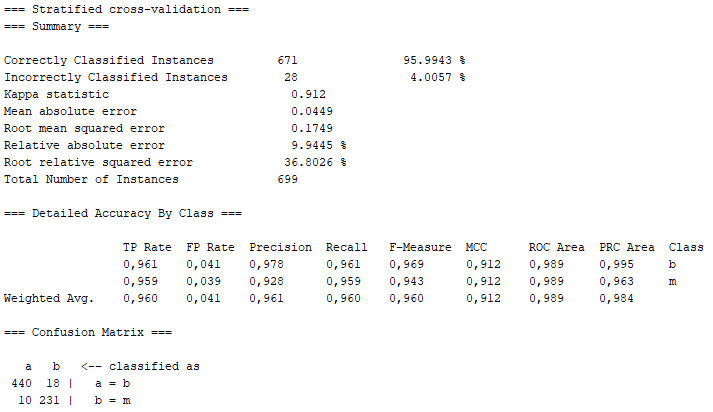
- 20 x 200 x 0.1:



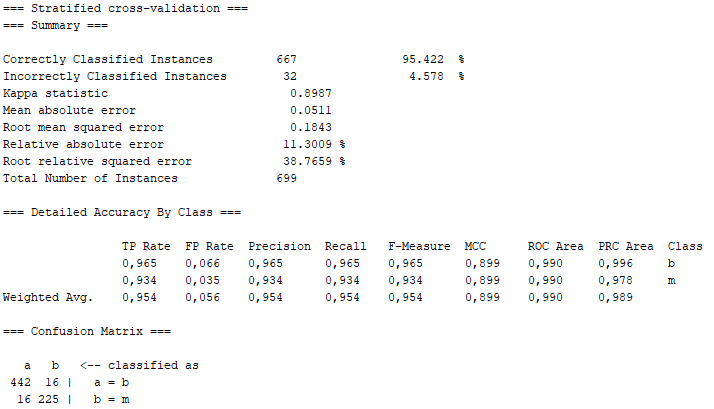
- 20 x 400 x 0.1:



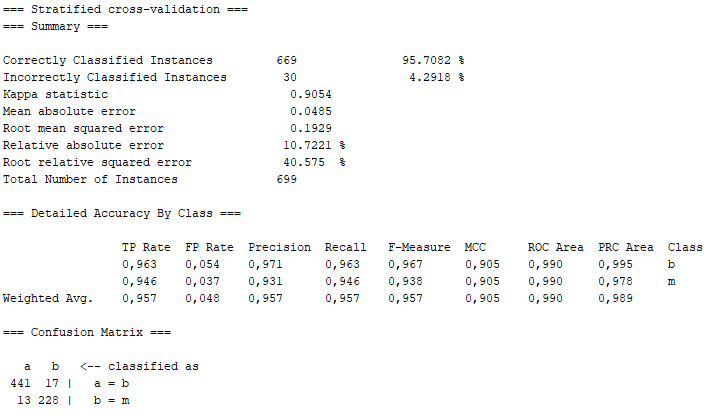
- 20 x 600 x 0.1:



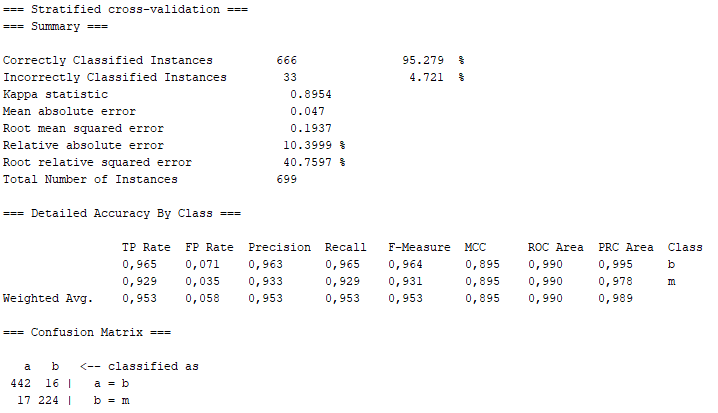
- 10 x 200 x 0.3:



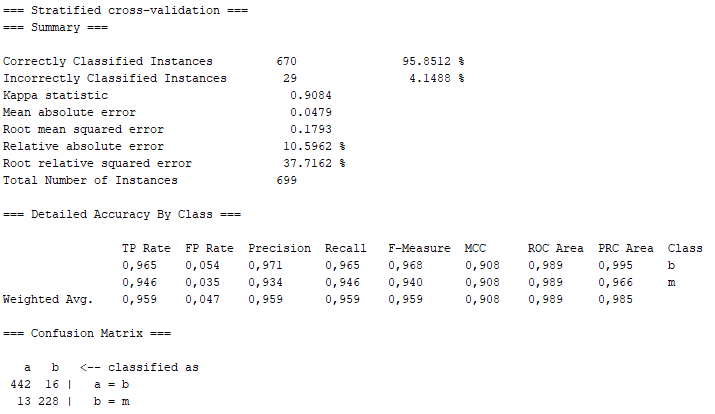
- 10 x 400 x 0.3:



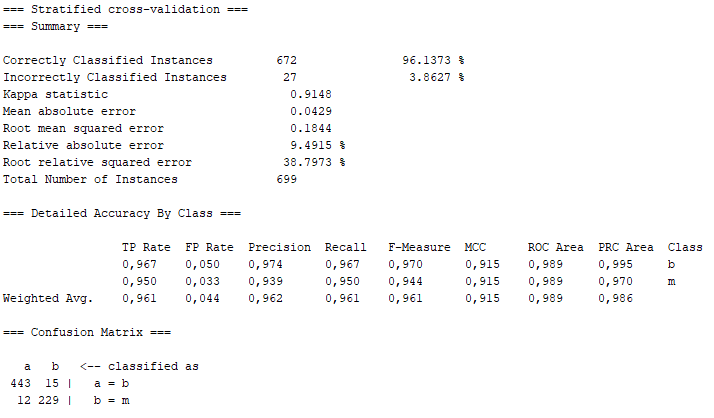
- 10 x 600 x 0.3:



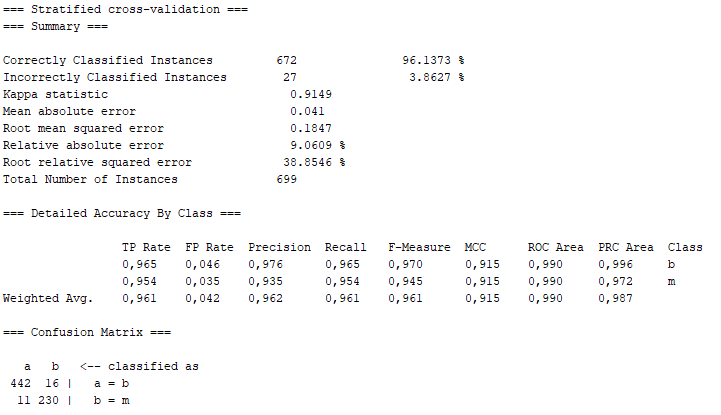
- 15 x 200 x 0.3:



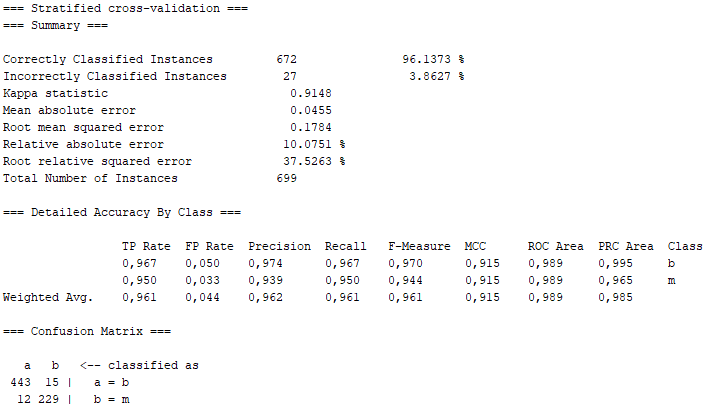
- 15 x 400 x 0.3:



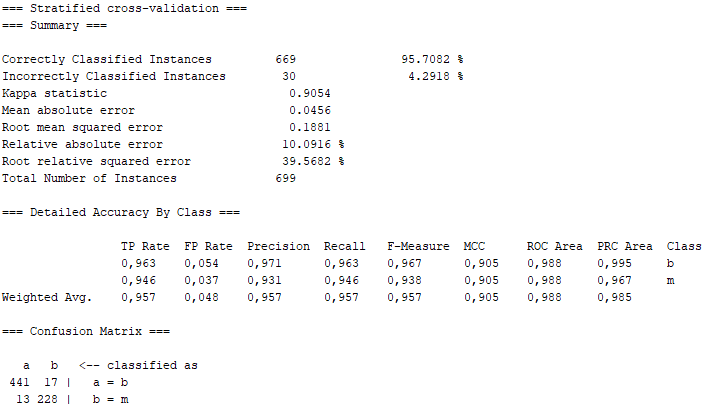
- 15 x 600 x 0.3:



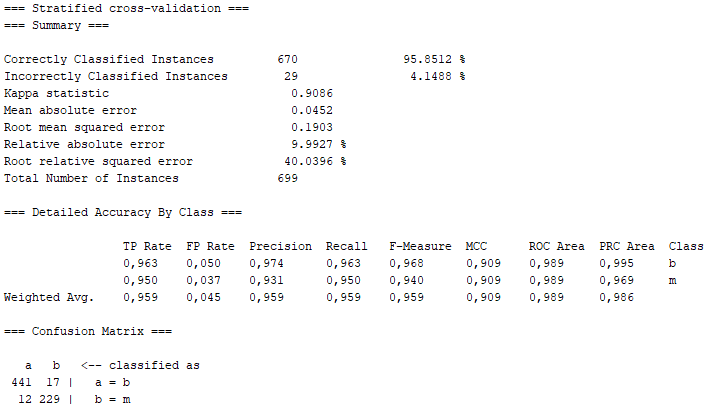
- 20 x 200 x 0.3:



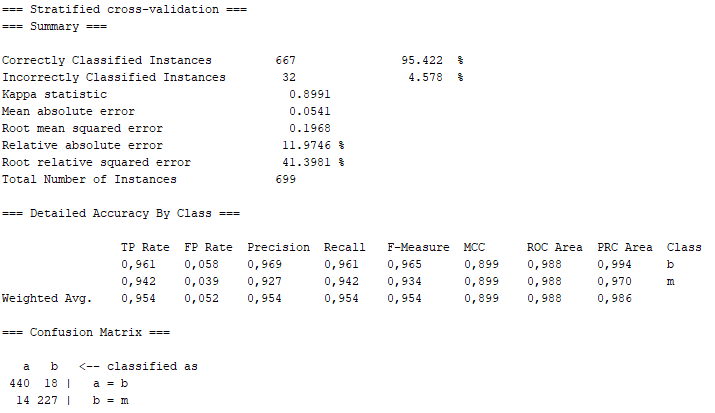
- 20 x 400 x 0.3:



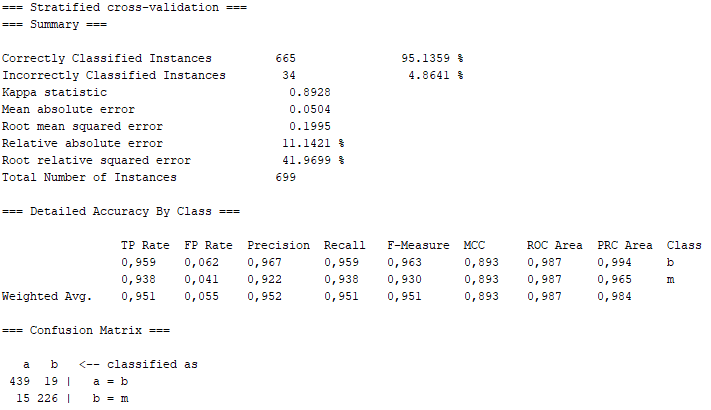
- 20 x 600 x 0.3:



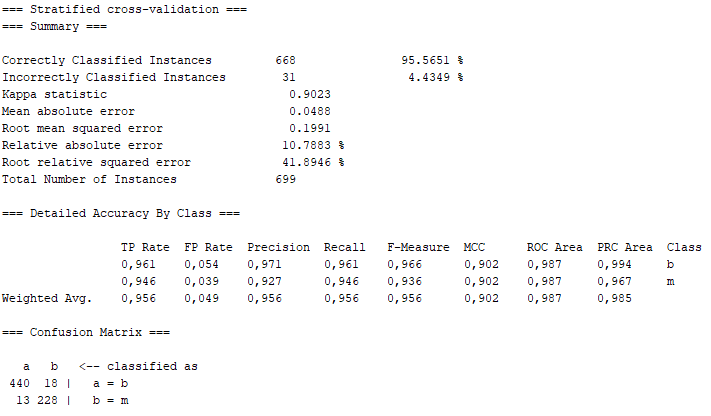
- 10 x 200 x 0.5:



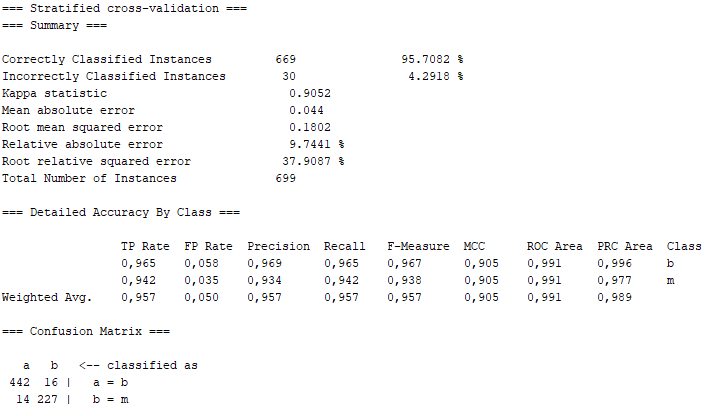
- 10 x 400 x 0.5:



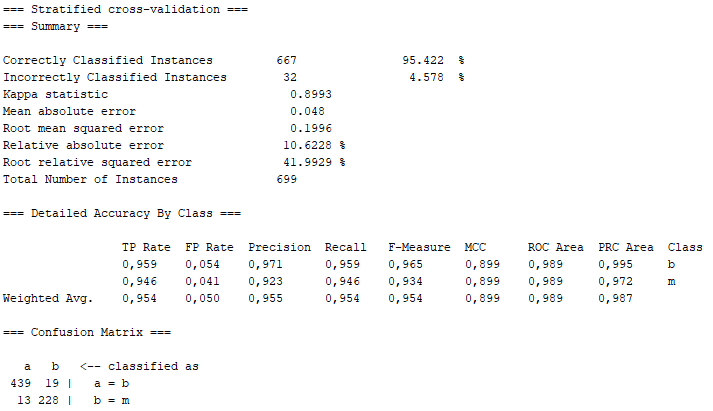
- 10 x 600 x 0.5:



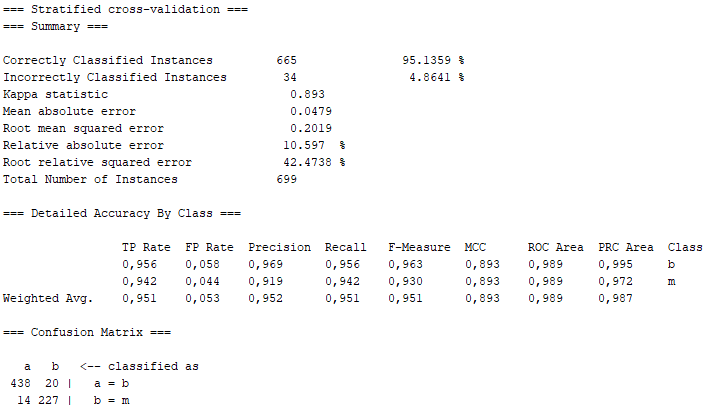
- 15 x 200 x 0.5:



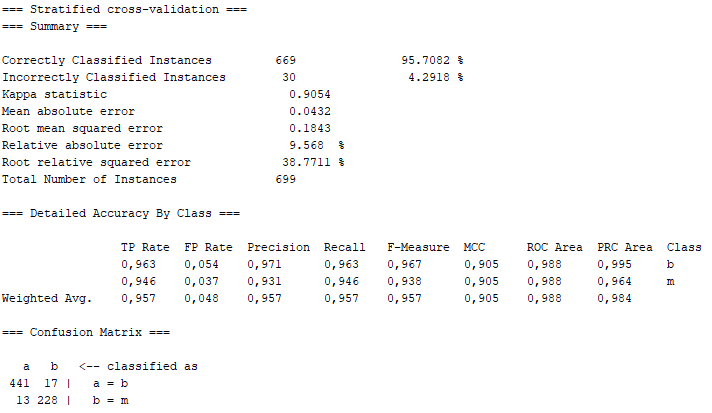
- 15 x 400 x 0.5:



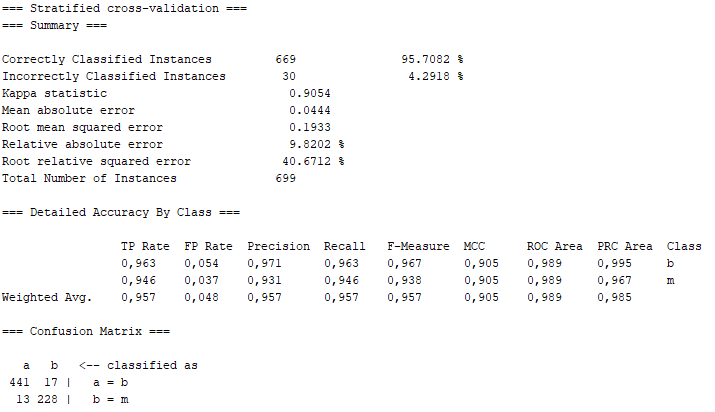
- 15 x 600 x 0.5:



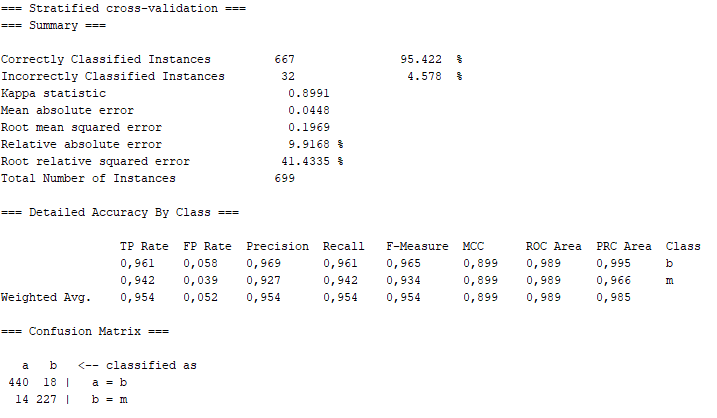
- 20 x 200 x 0.5:



- 20 x 400 x 0.5:



- 20 x 600 x 0.5:



O melhor desempenho foi com a Taxa de Aprendizagem de 0.3, com uma média de acurácia de 96,3916 %.

O melhor desempenho para cada um dos testes foi com as seguintes taxas de aprendizagem:

- 10 x 200: 0.3

- 10 x 400: 0.3

- 10 x 600: 0.3

- 15 x 200: 0.3

- 15 x 400: 0.3

- 15 x 600: 0.3

- 20 x 200: 0.3

- 20 x 400: 0.3

- 20 x 600: 0.3

**b) Qual rede teve a maior acurácia?**

5 configurações de rede apresentaram o maior valor de acurácia, sendo estas:

(Nº camadas intermediárias x Nº épocas x Taxa de aprendizagem : Acurácia)

- 15 x 200 x 0.3: 96,5665 %

- 15 x 400 x 0.3: 96,5665 %

- 15 x 600 x 0.3: 96,5665 %

- 20 x 400 x 0.3: 96,5665 %

- 20 x 600 x 0.3: 96,5665 %

**c) Qual o efeito do número de épocas sobre o treinamento, do número de nós na camada intermediária e da taxa de aprendizagem no desempenho da rede em termos de erro e acurácia?**

Um número alto de épocas pode aumentar a acurácia, pois fará o algoritmo processar os pesos por um período maior, porém se aumentarmos muito esse número, podemos chegar em um ponto onde a acurácia já estabilizou, ou seja, vai aumentar o processamento do treinamento sem melhorar a acurácia. Um número muito baixo de épocas pode não ser o suficiente para o algoritmo encontrar um peso bom, por exemplo, se a taxa de aprendizagem for baixa, o que faz os pesos serem alterados mais lentamente, diminuindo a acurácia.

Um número alto de nós na camada intermediária pode aumentar a acurácia, porém ele pode chegar em um ponto onde aumentamos o número de nós e a acurácia não aumenta, pois ela chegou em um ponto onde fica estabilizada, o que vai aumentar o processamento do treinamento desnecessariamente. Um número muito baixo de nós pode piorar o cálculo dos novos pesos, fazendo a acurácia diminuir.

A taxa de aprendizagem influencia na velocidade com que os pesos são alterados. Se ela for muito baixa, a precisão do ajuste de pesos aumenta, porém os pesos vão demorar muito para serem corrigidos para um valor ótimo, o que pode diminuir a acurácia ou exigir um número muito grande de épocas. Se ela for muito alta, os pesos sofrerão maior alteração, o que pode diminuir a precisão do ajuste de pesos e levar a uma situação onde o algoritmo nunca consegue encontrar um peso bom, o que pode diminuir muito a acurácia.

**d) Durante o treinamento é correto sempre utilizar uma rede com o maior número de nós possível nas camadas intermediárias?**

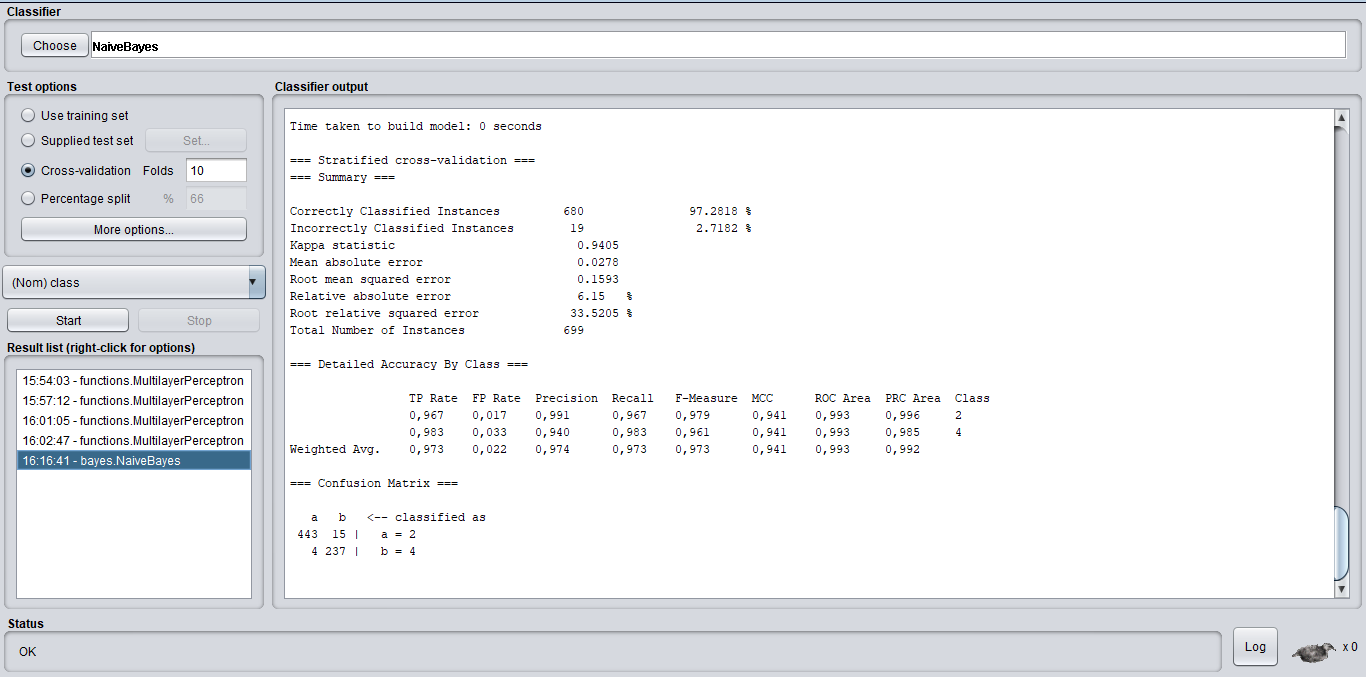
Não, pois aumentar muito o número de nós pode aumentar muito o processamento para o treinamento, sem aumentar a acurácia, visto que em um certo ponto a acurácia tende a estabilizar e parar de crescer.

**3) Compare o desempenho da melhor rede obtida no exercício 2 com aquele dos classificadores Naive Bayes e J48 (árvores de decisão) - use o WEKA.**

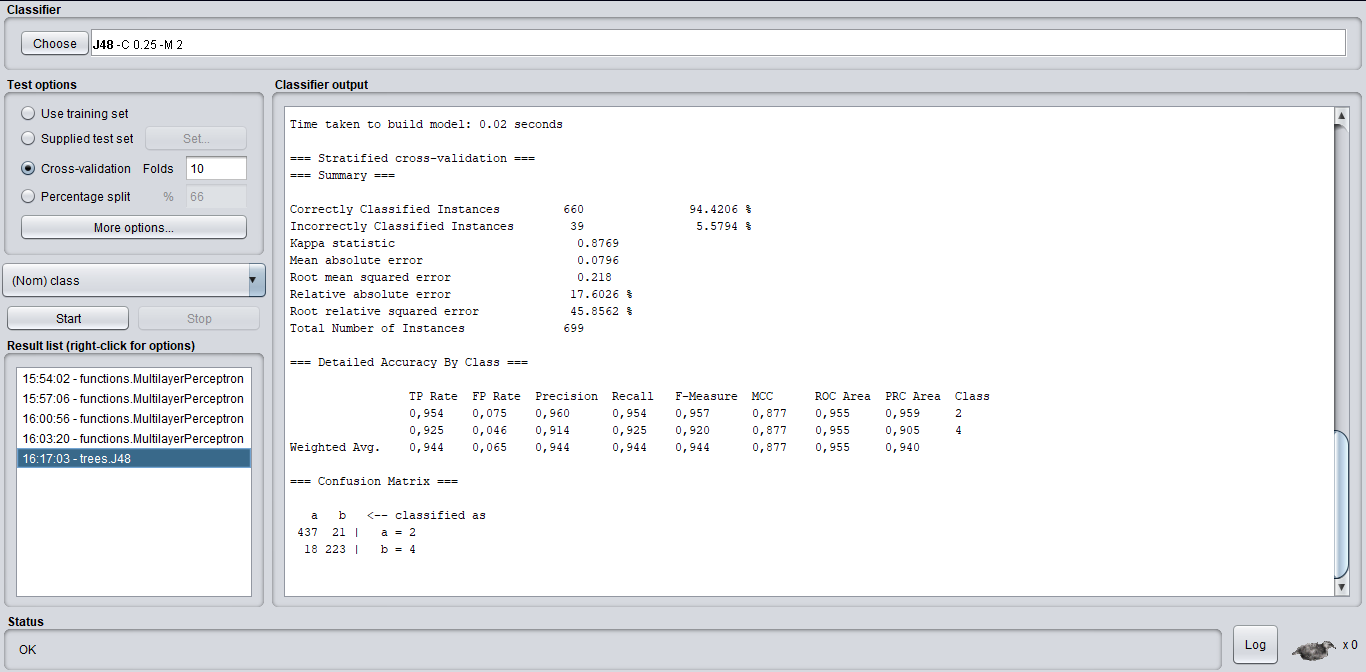
Melhor classificador do exercício 2:

Acurácia: 96,5665 %

Classificador Naive Bayes:



Classificador J48:



O melhor classificador foi o Naive Bayes, com acurácia de 97,2818 %, seguido da rede neural de múltiplas camadas, com 96,5665 %, e por último ficou o J48, com 94,4206 %.

**4) Leia a descrição do arquivo de dados semeion.csv no diretório UCI http://archive.ics.uci.edu/ml/datasets/Artificial+Characters. O arquivo sprimos.csv (anexo a esta atividade) foi gerado a partir daquele substituindo-se as saídas S1..S10 por uma única saísa S1. A saída do S1 classifica os números registrados nas imagems que deram origem ao arquivo como primo ou não primo. Utilize o software WEKA para treinar redes neurais para a classificação dos objetos deste arquivo. Durante os testes considere combinações de:**

**a) 3 diferentes taxas de aprendizado**

**b) 2 configurações da camada intermediária (máximo 15 nós)**

**d) 3 configurações do número de épocas (>200 e <1000)**

**Reporte os resultados e identifique o classificador com melhor desempenho.**

Taxas de aprendizagem escolhidas: 0,5 e 1 e 0,75

Camadas Intermediárias escolhidas: 6 e 12

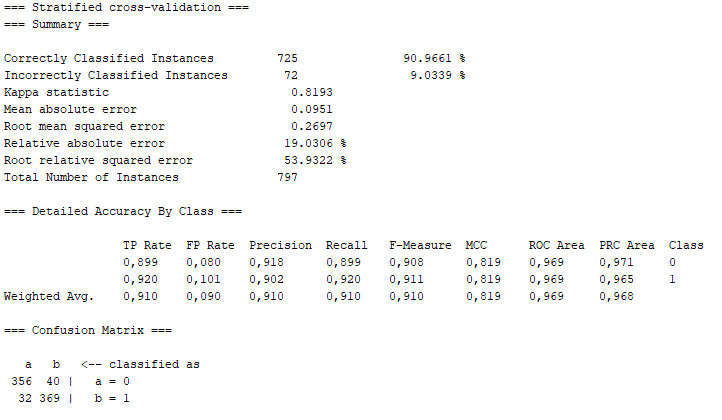
Número de Épocas escolhidas: 200 e 500 e 800

Resultados utilizando validação cruzada com 10 folds:

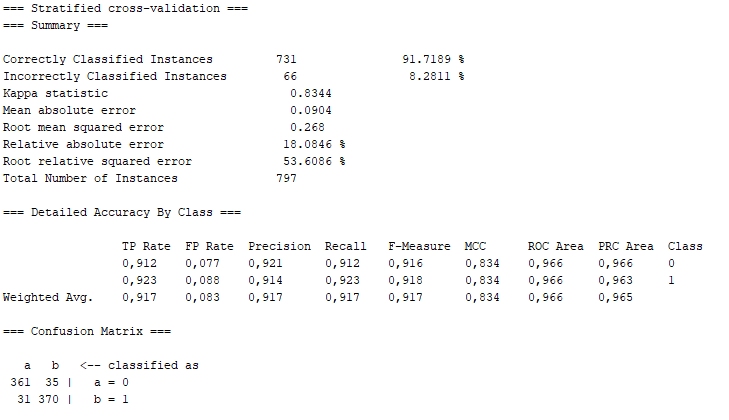
Os valores nos itens indicam, respectivamente:

- Taxa de Aprendizagem x Camadas Intermediárias x Número de Épocas:

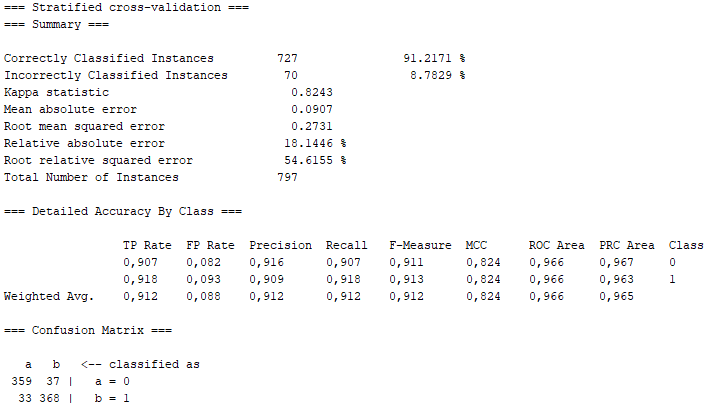
- 0,5 x 6 x 200:



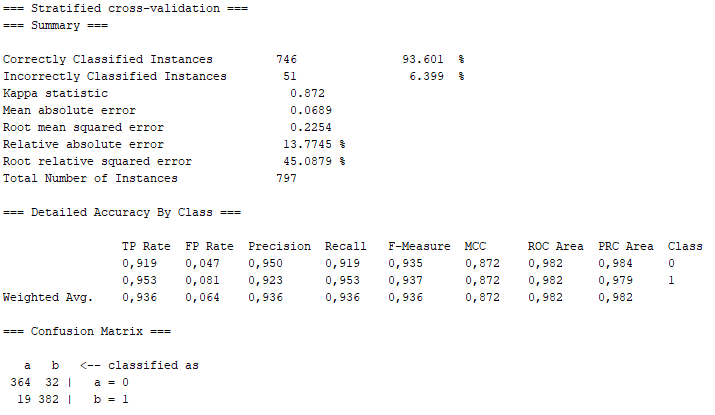
- 0,5 x 6 x 500:



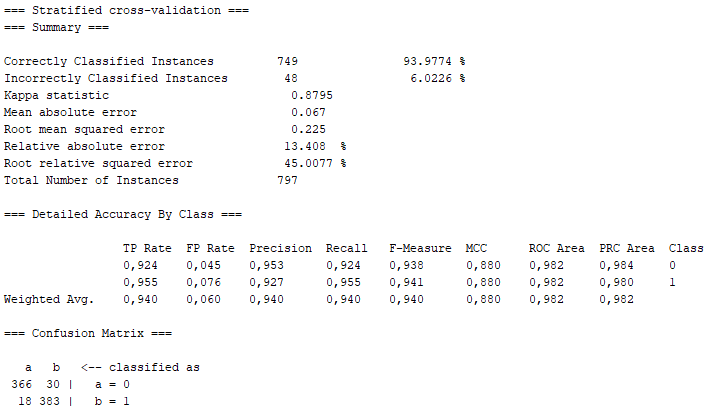
- 0,5 x 6 x 800:



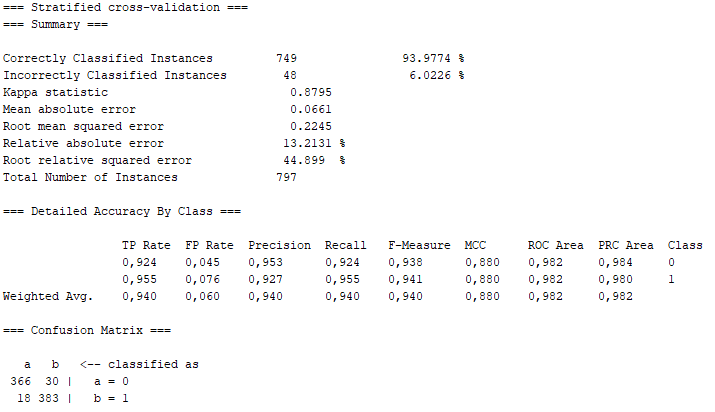
- 0,5 x 12 x 200:



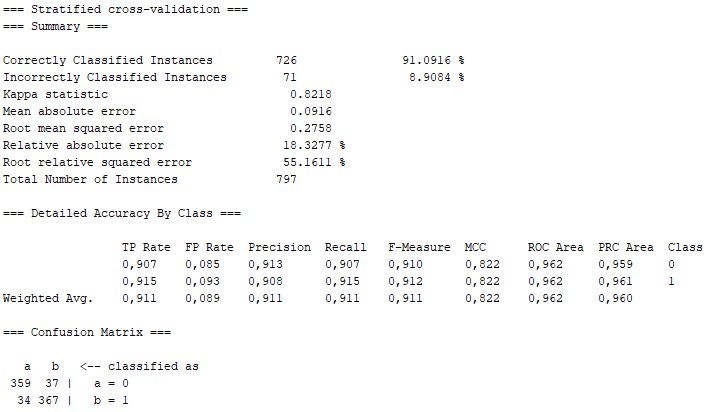
- 0,5 x 12 x 500:



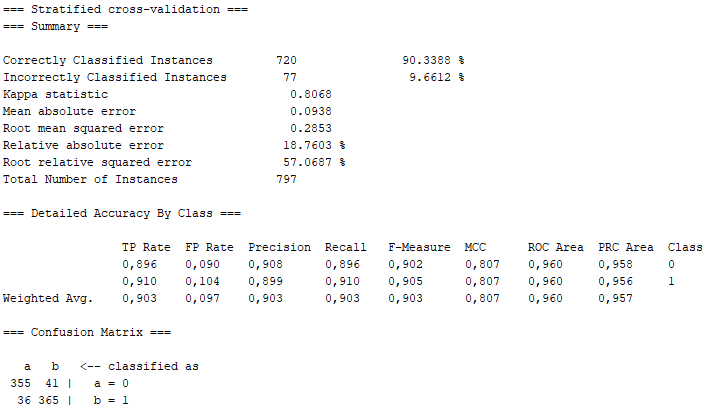
- 0,5 x 12 x 800:



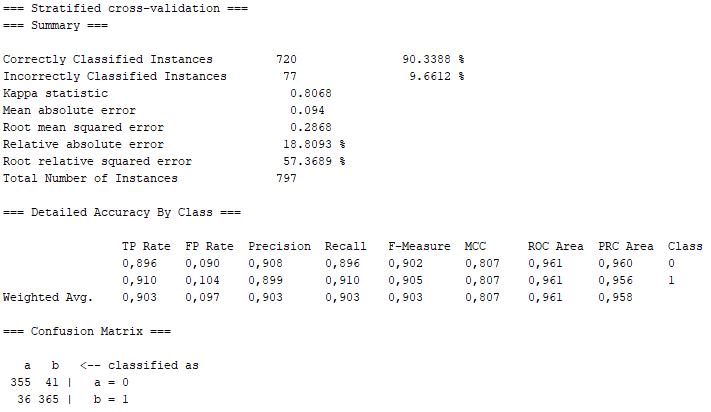
- 1 x 6 x 200:



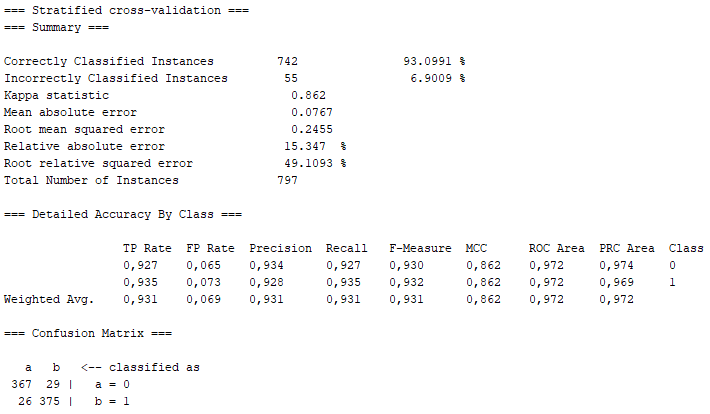
- 1 x 6 x 500:



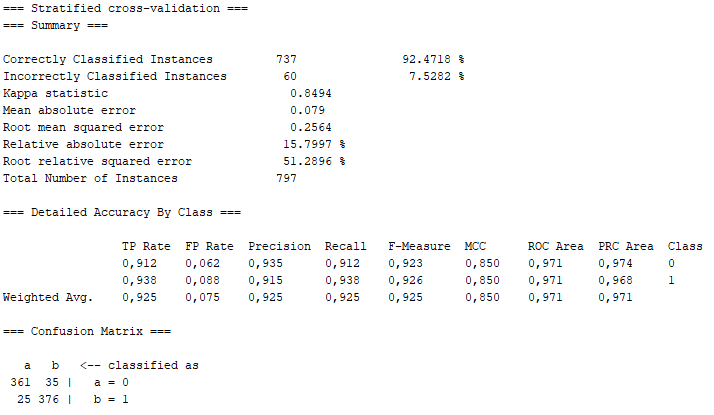
- 1 x 6 x 800:



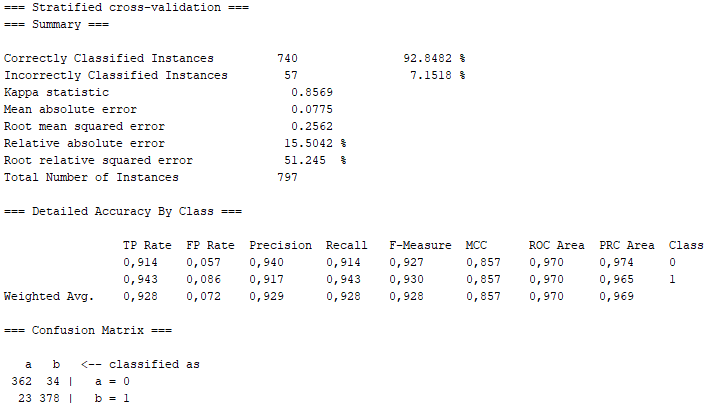
- 1 x 12 x 200:



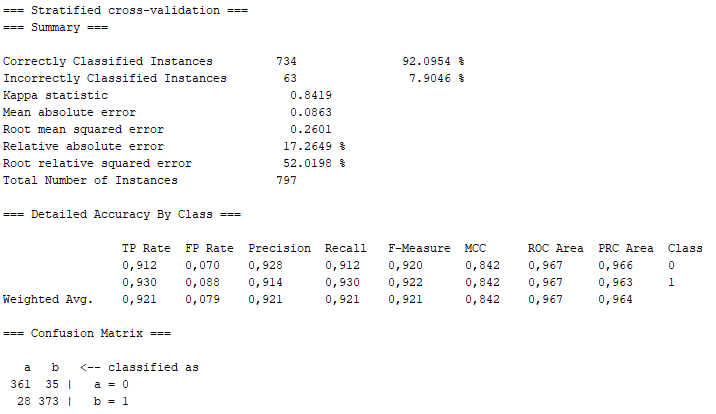
- 1 x 12 x 500:



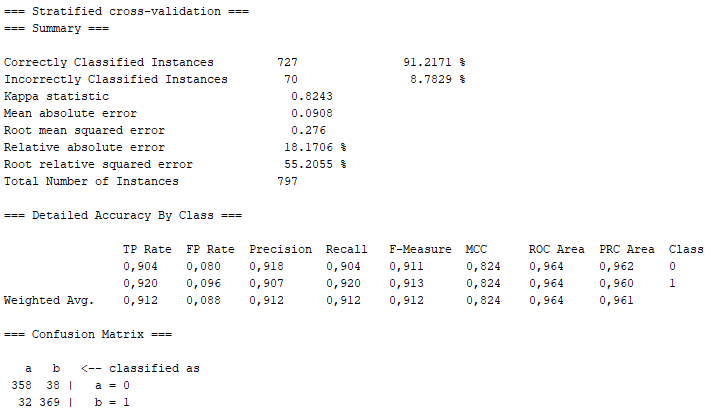
- 1 x 12 x 800:



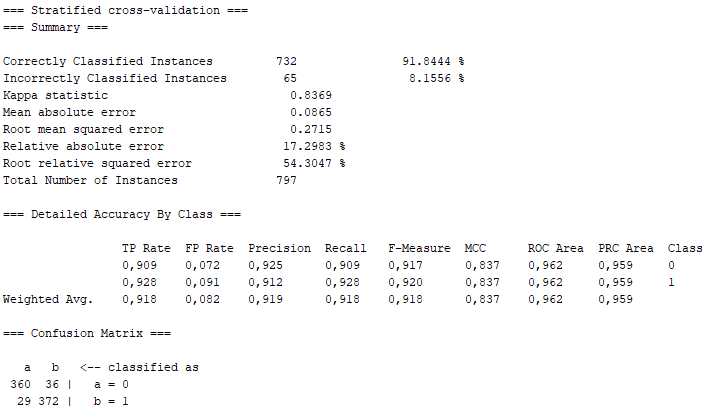
- 0,75 x 6 x 200:



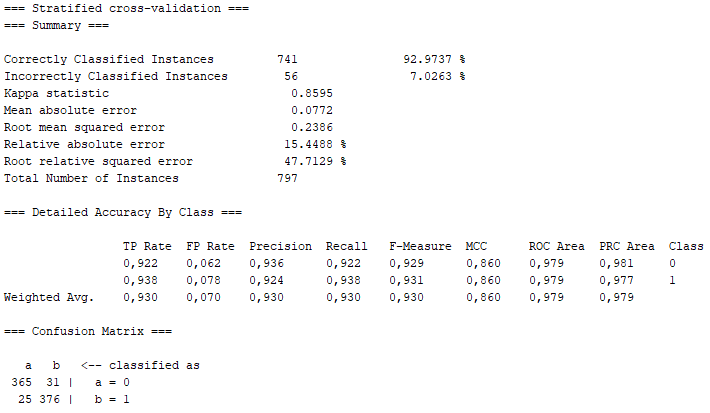
- 0,75 x 6 x 500:



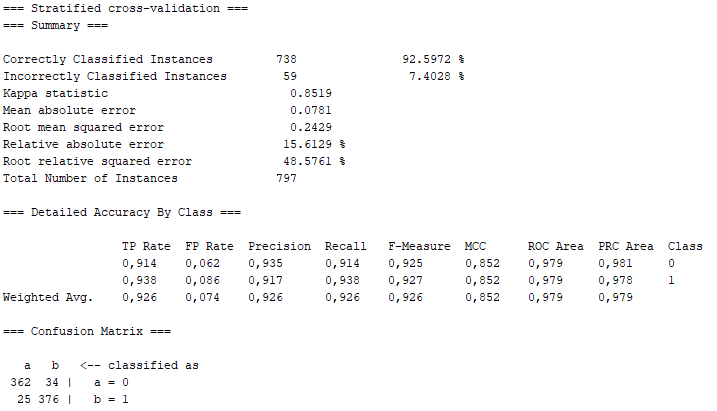
- 0,75 x 6 x 800:



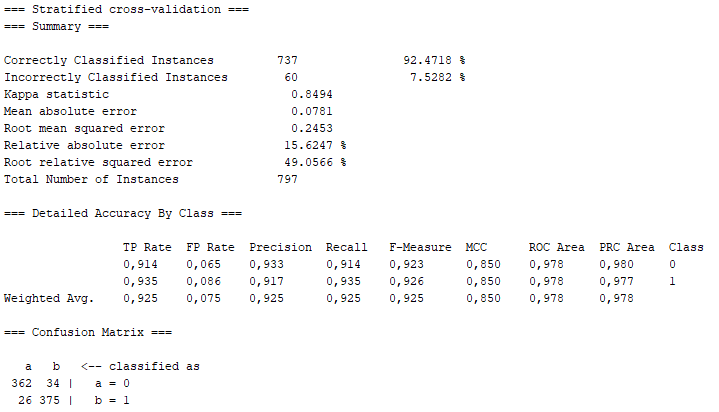
- 0,75 x 12 x 200:



- 0,75 x 12 x 500:



- 0,75 x 12 x 800:



Foi obtido dois classificadores com o melhor desempenho treinado, sendo estes:

- Taxa de Aprendizagem: 0.5

- Camadas Intermediárias: 12

- Número de Épocas: 500 e 800

A acurácia destes classificadores foi de 93,9774 %.