

Classificação de ruído em biosinais
Trabalho prático nº2, Computação Adaptativa

Adriano Vinhas (2009106560, avinhas@student.dei.uc.pt)
José Ribeiro (2008112181, jbaia@student.dei.uc.pt)

December 4, 2013

1 Introdução

Este trabalho, no âmbito da disciplina de Computação Adaptativa, tem como objectivo construir uma rede neuronal multimacada capaz de reconhecer se um sinal biológico deve ser descartado (devido à presença de ruído) ou deve ser analisado. Este reconhecimento é feito com base num conjunto de características do sinal que vão servir de entrada para a rede neuronal.

Este objectivo foi atingido fazendo um estudo paramétrico tendo em conta os seguintes parâmetros de estudo:

- Método de treino
- Coeficiente de aprendizagem
- Número de neurónios na camada escondida
- Dimensionalidade do problema
- Número de épocas do treino
- Aplicação de normalização sobre os dados

A parte que foi mais focada na realização deste trabalho foi o estudo paramétrico feito com base nos parâmetros acima indicados. Com base nos resultados obtidos, procurámos uma solução que nos permitisse chegar à combinação dos parâmetros que optimizasse os valores de sensibilidade e especificidade das saídas da rede.

2 Concepção da rede neuronal

Nesta secção estão descritas a arquitectura da rede usada para o trabalho e a forma como foi feito o treino da rede.

A arquitectura da rede neuronal usada para o trabalho prático caracteriza-se por 26 entradas e 1 saída, sendo que cada uma das entradas representa uma característica do *dataset* que nos foi fornecido e a saída representa a decisão (se o sinal deve ser descartado ou analisado). No entanto, para os testes efectuados, o número de entradas da rede está sujeito a alterações, sendo este número dependente do uso (ou não) de mecanismos de redução de dimensionalidade.

A função de activação usada para a camada interna foi a tangente hiperbólica, para garantir que a rede neuronal pudesse ser um aproximador universal, e para a camada de saída foi usada uma função linear.

A figura 1 representa a arquitectura da rede usada.

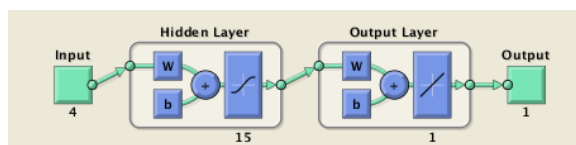


Figure 1: Arquitectura da rede usada (com 26 características e 15 neurónios na camada escondida)

O *dataset* usado para este trabalho é composto por um total de 12331 exemplos. Deste conjunto, 70% dos exemplos (8632) foram usados para treinar a rede, e os restantes 30% (3699) para validação desta e análise de resultados.

O método de selecção de exemplos para cada um dos conjuntos é aleatório. Para que tal aconteça, todos os exemplos são "baralhados" e ficam dispostos numa ordem completamente à ordem original. De seguida é feita a separação segundo a regra 70/30, tal como enunciado acima.

3 Estudo paramétrico

Nesta secção é explicada a metodologia de testes seguida para levar a cabo são apresentados os resultados do estudo paramétrico efectuado para obtenção do resultado óptimo encontrado.

Neste estudo, a ordem dos parâmetros testados torna-se fulcral para melhor explorar o espaço de combinações sem o fazer de forma exaustiva. No fim de cada variação paramétrica, o valor que optimizava as métricas usadas para medir a performance da rede neuronal era fixado na variação paramétrica seguinte. Por exemplo, após a determinação do número óptimo de neurónios para a camada escondida, esse valor era fixado para os restantes testes.

No caso do estudo do coeficiente de aprendizagem, tal não se aplica uma vez que este foi o primeiro parâmetro a ser variado.

Dado o impacto dos variados parâmetros consoante o método de treino, escolhemos efectuar todas as variações paramétricas fixando o método. Como tal, todos os testes relativos a `trainlm` foram executados separadamente dos testes relativos a `traingd`.

Para garantir significado estatístico nos resultados, as métricas de avaliação da rede neuronal (performance, sensibilidade e especificidade) apresentadas são a **média** de **30** execuções distintas.

A tabela 1 sumaria os parâmetros e com valores *default*.

Table 1: Parâmetros predefinidos usados para o estudo paramétrico

Parâmetro	Valor utilizado
Coeficiente de Aprendizagem	Não aplicável
Neurónios da Camada Escondida	5
Número de características	4
Número de épocas de treino	100
Normalização de input	Com

3.1 Coeficiente de aprendizagem

Uma vez que o treino de Levenberg-Marquardt usa um Coeficiente de Aprendizagem adaptativo, este parâmetro apenas se aplica ao método de treino do Gradiente Descendente.

Assim, para o `traingd`, foram usados os seguintes valores de Coeficientes de Aprendizagem: 0.001, 0.0001 e 0.00001, sendo os restantes valores atribuídos como previamente indicado.

Como tal, o melhor resultado registou-se para o valor de **0.0001**, onde se obteve uma **sensibilidade** de **79.4%** e uma **especificidade** de **78.0%**. Os restantes resultados foram descartados, uma vez que se obtiveram NaN para um dos valores (sensibilidade/especificidade), consequência de uma rede neuronal mal treinada (saída completamente igual a 0s ou 1s).

A tabela 2 sumaria os testes efectuados.

Table 2: Resultados obtidos no estudo do Coeficiente de Aprendizagem

	Performance	Sensibilidade	Especificidade
traingd: 0.001	49.31%	NaN	NaN
traingd: 0.0001	78.27%	79.43%	78.02%
traingd: 0.00001	57.96%	NaN	57.96%

3.2 Número de Neurónios na Camada Escondida

Para todas as variações do número de Neurónios da Camada Escondida foi usado o valor de Coeficiente de Aprendizagem = **0.0001** e os restantes parâmetros atribuídos conforme a tabela 1. De realçar também que para cada valor deste parâmetro ambos os métodos de treino foram testados.

Foram usados os seguintes valores para o Número de Neurónios da Camada Escondida: 2, 5, 8 e 10.

Os melhores resultados foram obtidos para 5 Neurónios com **trainlm**, com uma **sensibilidade** de **82.4%** e uma **especificidade** de **88.4%**.

Comparando os métodos de treino, salta à vista o facto de, em todos os casos, o treino de **Levenberg-Marquardt** obter melhores resultados quando comparado ao método de treino do Gradiente Descendente.

Em relação à escolha do número de neurónios escolhemos o valor de **5** neurónios porque nos pareceu que era o número de neurónios que garantia o compromisso entre a capacidade de resolução do problema e a complexidade da solução. No caso dos 2 neurónios, os valores de todas as métricas são mais baixos, demonstrando assim a incapacidade da rede de aproximar o comportamento. Para os valores de 8 e 10 neurónios, o rácio sensibilidade-especificidade é mais desequilibrado, o que indicia que o número de neurónios escolhidos é inadequado porque a rede perde capacidade de generalização.

A tabela 3 resume os testes efectuados.

Table 3: Resultados obtidos no estudo do número de Neurónios da Camada Escondida

	Performance	Sensibilidade	Especificidade
traingd: 2	77.29%	78.77%	76.81%
traingd: 5	78.27%	79.43%	78.02%
traingd: 8	77.99%	79.18%	77.78%
traingd: 10	78.09%	79.24%	78.17%
trainlm: 2	84.86%	81.31%	87.47%
trainlm: 5	85.92%	82.43%	88.42%
trainlm: 8	85.49%	80.80%	89.11%
trainlm: 10	85.21%	80.45%	88.73%

3.3 Dimensionalidade do Problema

Para estudar a variação deste parâmetro foi usado o valor de Coeficiente de Aprendizagem = **0.0001** e um Número de Neurónios = **5** (melhores valores até agora obtidos). Para reduzir a dimensionalidade do problema escolhemos utilizar *Principal Component Analysis* (PCA). O número de características cuja variância cumulativa perfaz um valor superior a 95% foi de **4**. No entanto, quisemos testar também os valores de 2 e 6. Como esperado, o valor de **6** obteve os melhores valores de sensibilidade e especificidade; no entanto, o ganho relativo a 4 características não é superior ao ganho obtido entre 4 características e 2. Isto leva-nos a concluir que a adição de novas características às entradas da rede não acrescentam a mesma quantidade de informação.

Table 4: Resultados obtidos no estudo do número de Características

	Performance	Sensibilidade	Especificidade
traingd: 2	76.74%	79.69%	75.82%
traingd: 4	78.09%	79.24%	78.17%
traingd: 6	80.79%	83.52%	80.20%
trainlm: 2	84.04%	90.81%	40.49%
trainlm: 4	85.21%	80.45%	88.74%
trainlm: 6	86.60%	83.44%	83.84%

3.4 Número de Épocas de Treino

Para testar os diferentes números de Épocas de Treino foi usado o valor de Coeficiente de Aprendizagem = **0.0001**, um Número de Neurónios = **5** e também usámos 4 e 6 características. No entanto, os resultados exibidos apenas cobrem o caso das 6 características. Em relação à variação deste parâmetro, testámos os valores de 30, 50 e 100 épocas de treino. Como esperado, o valor de **100** épocas mostrou-se o mais adequado (para qualquer um dos métodos de treino) para a convergência da rede neuronal para uma melhor solução. Testes para 200 épocas demonstraram um ganho não significativo podendo indiciar *overfitting*.

Table 5: Resultados obtidos no estudo do número de Épocas de Treino

	Performance	Sensibilidade	Especificidade
traingd: 30	60.94%	NaN	60.58%
traingd: 50	66.66%	NaN	65.21%
traingd: 100	78.27%	79.43%	78.02%
trainlm: 30	84.85%	80.29%	88.31%
trainlm: 50	85.08%	80.30%	88.63%
trainlm: 100	85.92%	82.43%	88.42%

3.5 Normalização dos dados

Esta subsecção pretende avaliar a influência da normalização dos dados na qualidade da rede neuronal. Uma vez que todos os testes foram efectuados com os dados normalizados, bastou-nos verificar como a performance, sensibilidade e especificidade das melhores redes neuronais para cada método variavam quando os dados eram **não normalizados**. Como esperado, a qualidade da rede neuronal foi **inferior** para ambos os testes.

Table 6: Resultados obtidos no estudo da Normalização dos Dados

	Performance	Sensibilidade	Especificidade
traingd : Com	78.27%	79.43%	78.02%
traingd : Sem	64.59%	87.81%	62.86%
trainlm : Com	85.92%	82.43%	88.42%
trainlm : Sem	78.24%	84.71%	46.36%

3.6 Método de treino

Para chegar à conclusão de qual o melhor método de treino a usar para treinar a rede neuronal foram usados os métodos de treino **trainlm** e **traingd**. Para cada um destes métodos de treino foram feitas as variações explicadas nas subsecções anteriores, a cada um dos parâmetros. No fim, apenas se atingiu a conclusão do melhor método de treino escolhendo o método correspondente à solução com maiores valores de sensibilidade e especificidade.

Por consequência das subsecções anteriores chegou-se à conclusão que o método de treino de **Levenberg-Marquardt** atingiu sempre valores superiores de performance, sensibilidade e especificidade, quando comparado com o método de treino do Gradiente Descendente. Por este motivo, a nossa melhor rede neuronal teve como método de treino a primeira opção.

4 Conclusão

Depois do estudo paramétrico efectuado a rede neuronal que obteve melhores resultados tinha um valor de sensibilidade de **83.44%** e de especificidade igual a **83.84%**.

Atingimos estes valores usando a seguinte parametrização:

- Método de treino: `trainlm` (Levenberg-Marquardt)
- Coeficiente de aprendizagem: Não aplicável
- Número de neurónios na camada escondida: 5
- Dimensionalidade do problema: 6
- Número de épocas do treino: 100
- Aplicação de normalização sobre os dados: Sim

Os valores obtidos de performance, sensibilidade e especificidade (em gráfico de barras) estão exibidos na figura 2.

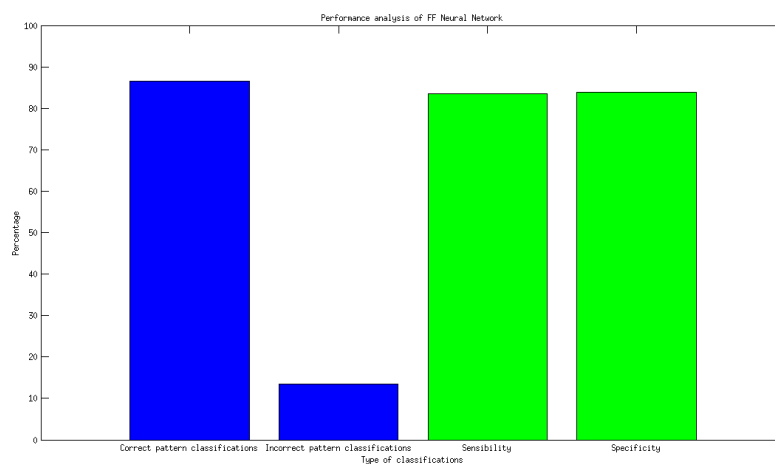


Figure 2: Performance, Sensibilidade e Especificidade da melhor rede encontrada