

Instituto Superior Técnico

Departamento de Engenharia Electrotécnica e de Computadores Machine Learning

3 rd Lab Assignment

Shift: 17:00 Terças Group Number: 19

Number:78308 Name Rui Daniel Ribeiro Dias

Number 75494 Name Nuno Pereira Azevedo Wallestein Teixeira

Multilayer perceptrons

This assignment aims at illustrating the applications of neural networks. In the first part we'll train a multilayer perceptron (MLP) for classification and in the second part we will train a MLP for regression.

This assignment requires MatLab's Neural Network Toolbox.

1 Classification

Our classification problem is a pattern recognition one, using supervised learning. Our goal is to classify binary images of the digits from 0 to 9, with 5x5 pixels each. The following figure illustrates some of the digits.

1) Try to find a parameter set (step size and momentum) that approximately minimizes the number of training epochs of the network. Indicate the values that you obtained.

Step size : 10

Momentum:0.6

Com estes valores obtivemos 31 iterações.

2) Determine how many epochs it takes for the desired minimum error to be reached (execute at least five tests and compute the median of the numbers of epochs).

1º32

2º33

3º34

4º35

5º63

Mediana: 34

1.4. Gradient method with adaptive step sizes and momentum

1) Train the network using the same initial values of the step size and momentum parameters as before. How many epochs are required to reach the desired minimum error? Make at least five tests and compute the median of the numbers of epochs.

1º 112

2º 109

3º 113

4º 107

5º108

Mediana: 109

É possível verificar que usando o método de gradiente com adaptive step sizes, a rede não é tão sensível à variação dos valores da learning rate e momentum constante comparativamente com o step fixo. Isto acontece uma vez que ele adapta o step size a cada iteração, sendo desse modo insensível aos valores iniciais.

2) Try to approximately find the set of parameters (initial step size and momentum) which minimizes the number of training epochs. Indicate the results that you have obtained (parameter values and median of the numbers of epochs for training). Comment on the sensitivity of the number of training epochs with respect to variations in the parameters, in comparison with the use of a fixed step size parameter. Indicate the main results that led you to your conclusions.

Step size: 0.8 Momentum: 0.6

Iterações : 54,70,71,74,82 Mediana: 71

Com um passo fixo a variação é muito mais acentuada do que quando usamos um adaptive step size. Isto acontece uma vez que no adaptive step size, este adapta-se a cada iteração, fazendo com que mesmo que o passo inicial não seja o mais apropriado para encontrar o mínimo, o próprio algoritmo adapta-se fazendo com que o número de interações necessárias não tenha uma dependência considerável com os parâmetros iniciais.

Comment on the values in the confusion matrix you obtained. Is the accuracy the same for every digit? Are the confusions between the different digits what you'd expect?

Confusion Matrix										
Output Class	1	2	3	4	5	6	7	8	9	
	5 3.1%	0 0.0%	0 0.0%	0 0.0%	1 0.6%	0 0.0%	1 0.6%	0 0.0%	2 1.2%	0 0.0%
	1 0.6%	16 10.0%	0 0.0%	1 0.6%	0 0.0%	0 0.0%	1 0.6%	3 1.9%	0 0.0%	0 0.0%
	2 1.2%	0 0.0%	9 5.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	0 0.0%	0 0.0%	0 0.0%	9 5.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.6%	0 0.0%
	1 0.6%	0 0.0%	0 0.0%	0 0.0%	14 8.8%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%
	1 0.6%	0 0.0%	2 1.2%	2 1.2%	0 0.0%	14 8.8%	5 3.1%	0 0.0%	3 1.9%	0 0.0%
	3 1.9%	0 0.0%	2 1.2%	1 0.6%	1 0.6%	0 0.0%	9 5.6%	0 0.0%	1 0.6%	0 0.0%
	1 0.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	13 8.1%	0 0.0%	0 0.0%
	2 1.2%	0 0.0%	3 1.9%	3 1.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	8 5.0%	0 0.0%
	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	2 1.2%	0 0.0%	0 0.0%	1 0.6%	16 10.0%
	31.2% 68.8%	100% 0.0%	56.2% 43.8%	56.2% 43.8%	87.5% 12.5%	87.5% 12.5%	56.2% 43.8%	81.2% 18.8%	50.0% 50.0%	100% 0.0%
Target Class										70.6% 29.4%

É possível perceber que, nem todos os dígitos possuem o mesmo grau de previsão. É fácil perceber que há números mais fáceis de identificar do que outros, dadas as suas características únicas, enquanto que outros possuem uma estrutura semelhante à de outros números e por isso podem gerar mais conclusões erradas.

Write down the performance values that you obtained:

Training error (mse): 0.0498 Train set Accuracy: 82.0%
 Testing error (mse): 0.05924 Test set Accuracy: 70.6%

Which quantity (mean squared error, or global percentage of correct classifications) is best to evaluate the quality of networks, after training, in this problem? Why?

Global percentage, o MSE pode não ter os dados normalizados.

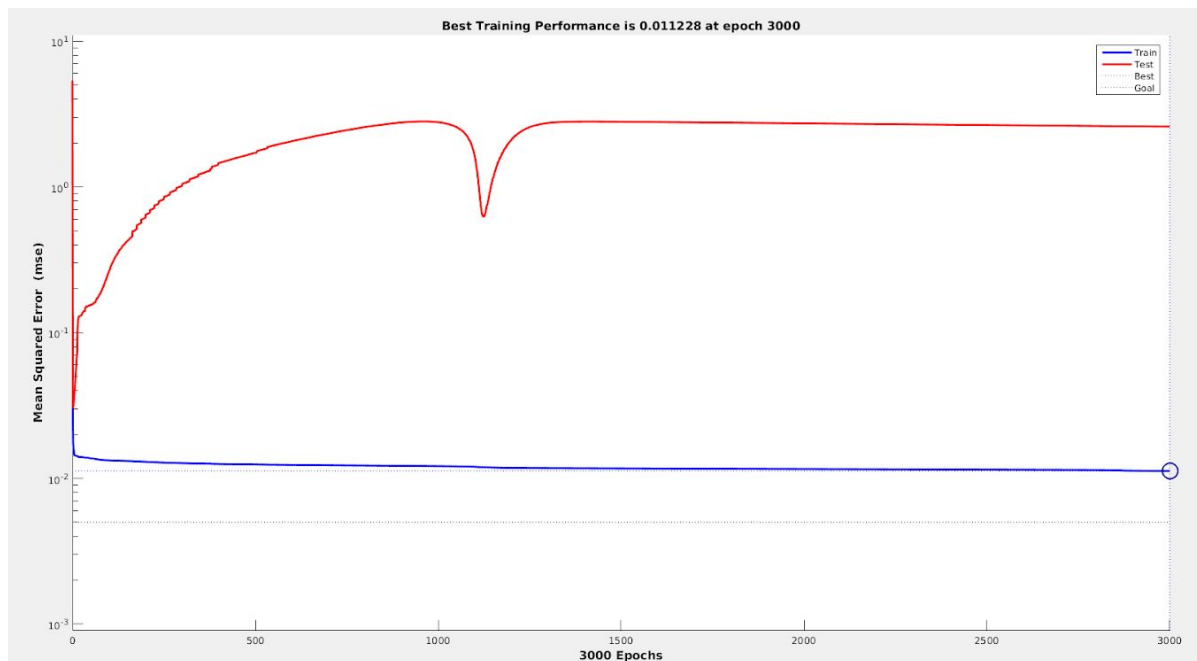
Global percentage é a melhor quantidade para avaliar a qualidade das networks por duas razões. Uma delas porque o EMQ não nos dá uma medida normalizada e portanto não é possível perceber se o método é bom ou não de um modo genérico apenas olhando para esta quantidade. Outra razão é considerando que existem erros localizados bastante elevados, o método vai avaliar um erro quadrático grande mas face à quantidade de dados que temos, o modelo pode fazer boas previsões, portanto esta quantidade não é tão relevante para avaliar a qualidade da rede como a percentagem global, contudo as duas quantidades estão relacionadas.

2. Regression

The goal of this second part is to estimate a function and to illustrate the use of a validation Set.

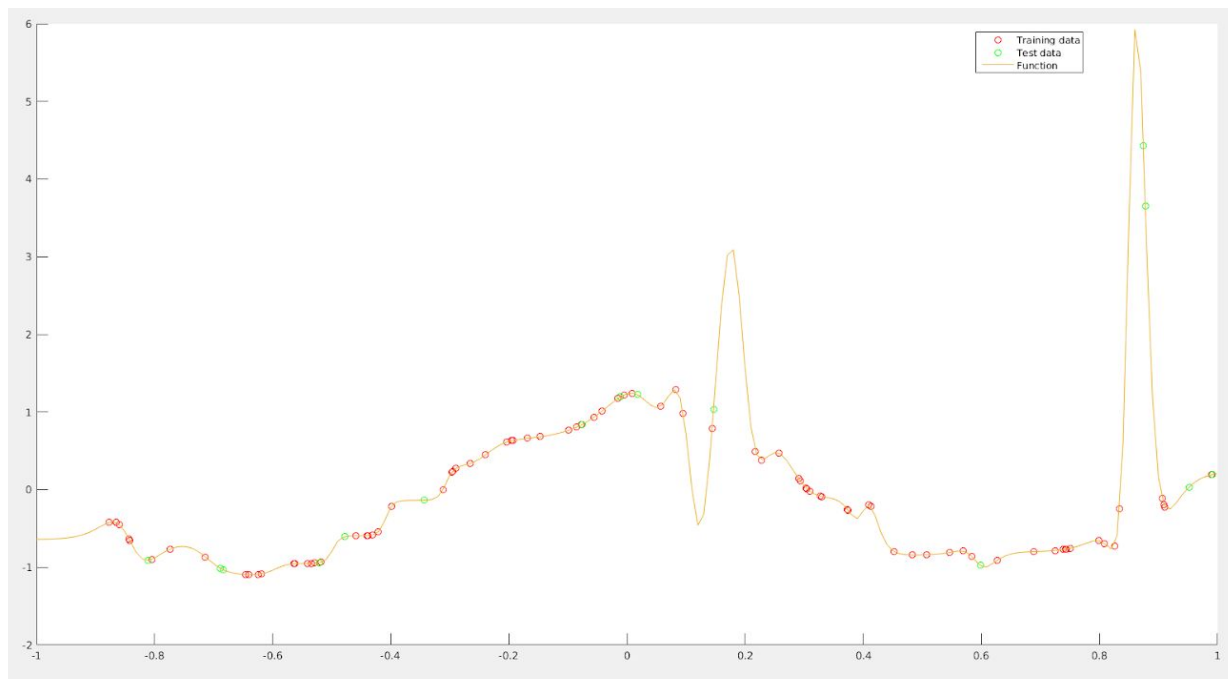
2.3 Training without a validation set

Train the network using the first 85 patterns for training and the last 15 for testing. Click the "Performance" button. Observe the evolution of the cost function for the different sets and comment.



No que toca a relação entre a test data, training data e o erro quadrático mínimo, podemos ver na figura acima que o EQM da test data, atinge um mínimo na primeira iteração e aumenta rapidamente para o resto das épocas. Isto é um sinal de overfitting no sentido que a nossa rede neuronal está a adquirir mais informação sobre o ruído da training data que é diferente da test data.

Obtain the estimated function for between -1 and 1 with a step of 0.01. Plot the training data, the test data and the estimated function, all in the same figure, and comment.



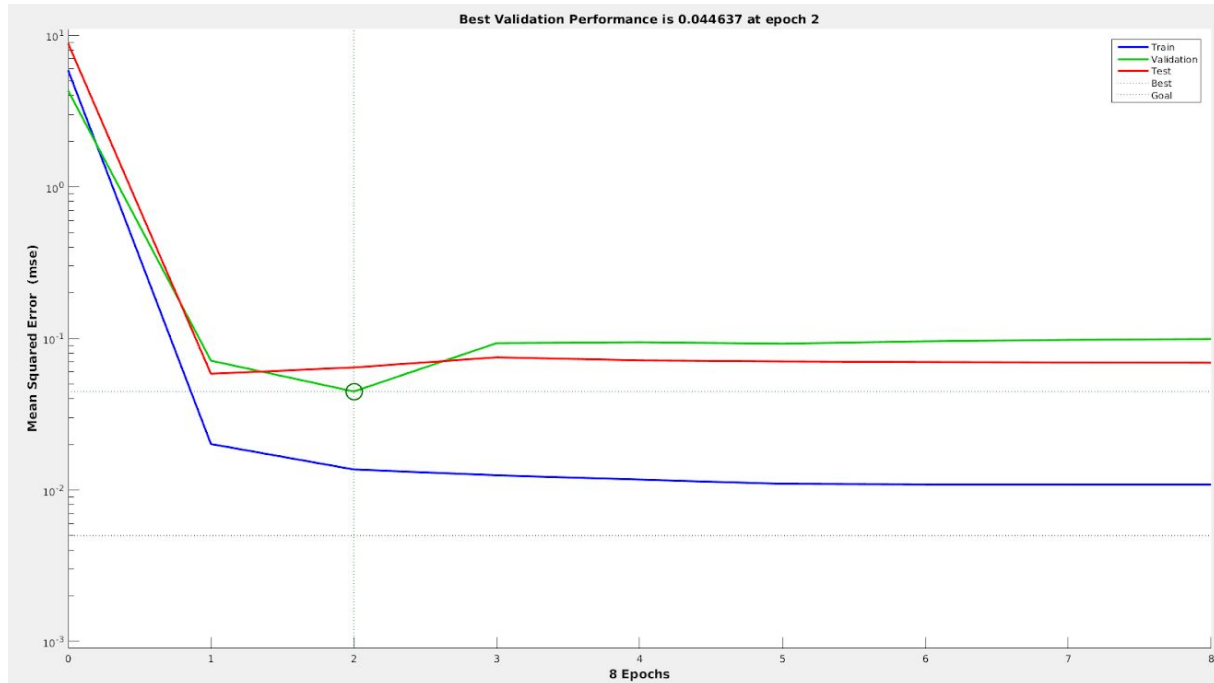
Neste gráfico podemos ver um exemplo de overfitting devido ao facto de não estarmos a considerar os pesos relativos ao mínimo do EQM face à época, o que produz um modelo bias face aos dados antigos(Overfitting)

2.4 Training with a validation set

When performing training with a validation set, how are the weights that correspond to the result of the training process chosen? What is the goal?

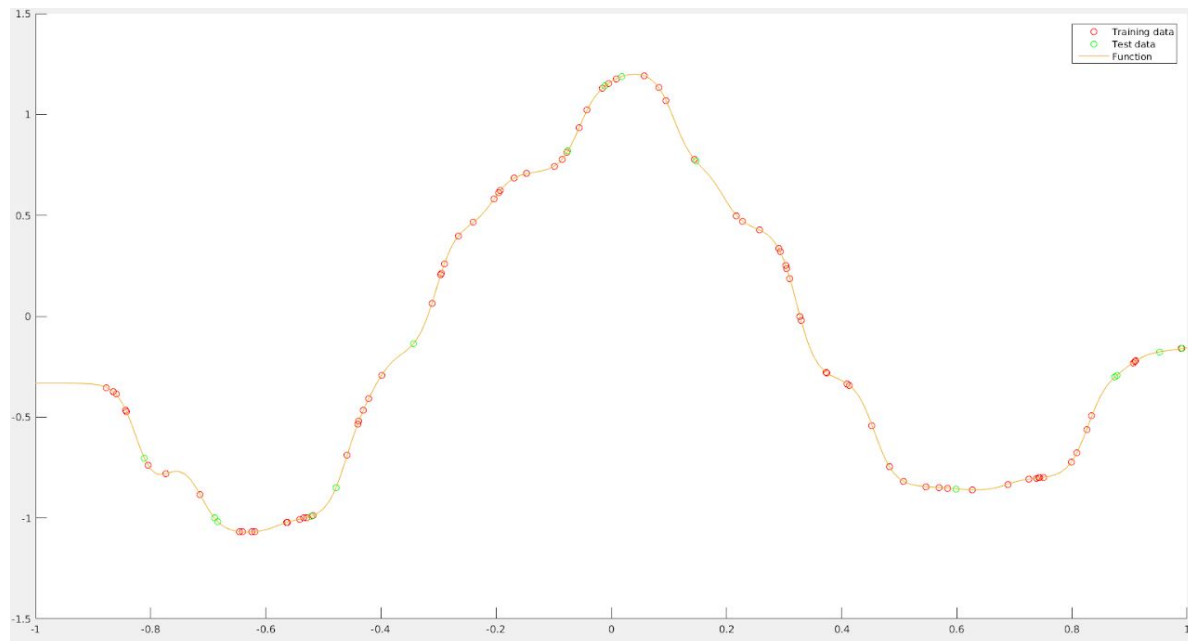
Quando as redes neurais são treinadas, sem restrições no número de épocas, o erro quadrático mínimo inicialmente diminui, devido à aprendizagem do método e depois aumenta devido ao overfitting. O objectivo de treinar a rede neuronal com o validation set é escolher os pesos optimos a partir do qual o erro quadrático médio atinge um mínimo face à iteração da época, como é possível ver na questão seguinte.

Train the network using the first 70 patterns for training, the next 15 for validation and the last 15 for testing. Click the "Performance" button. Comment on what is shown in the plot.



Neste gráfico conseguimos ver a eficácia de usar uma training data com validation set. Mal o erro quadrático atinja um mínimo, o modelo pára de iterar a época e escolhe os pesos ótimos nessa iteração.

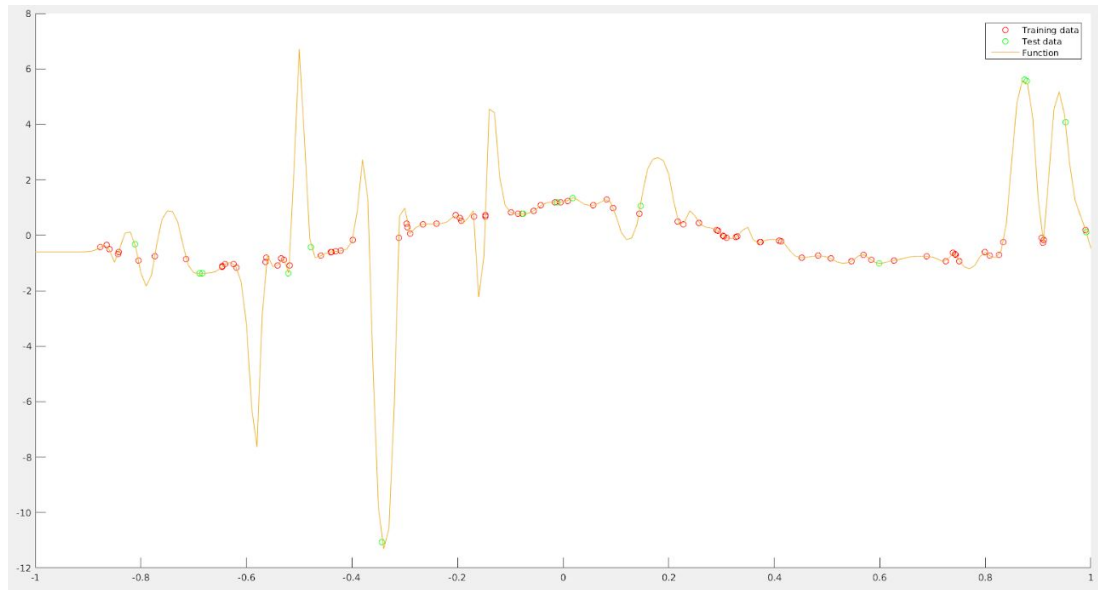
Obtain the newly estimated function and plot it in the same figure as the function obtained in 2.3. Comment



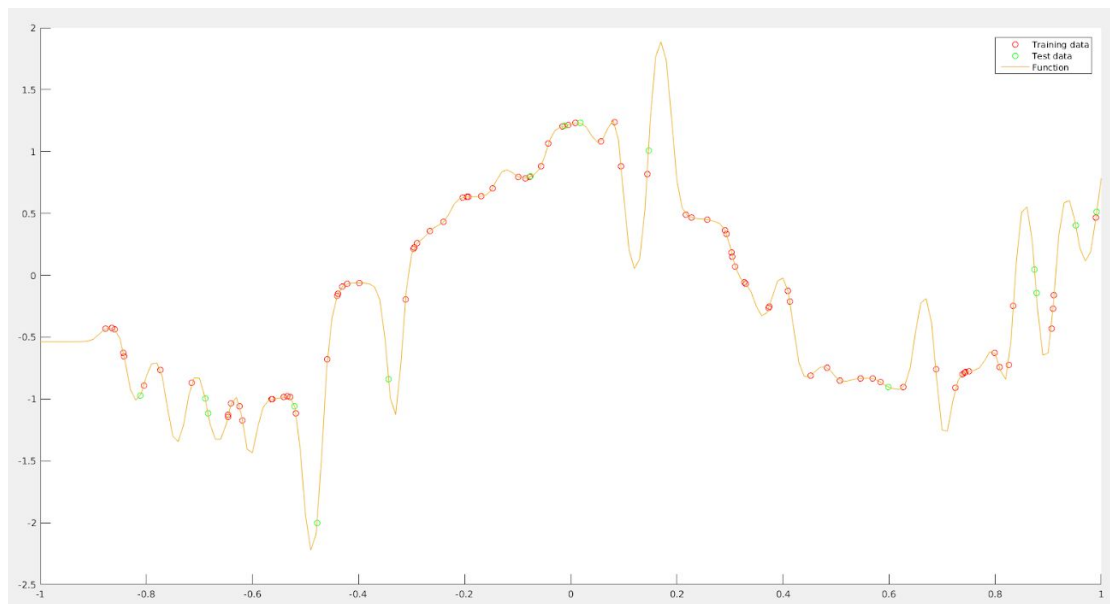
Este gráfico , demonstra o resultado de usar training data com validation data. Face aos dados, podemos ver que a nossa rede neuronal adapta-se muito bem face aos dados escolhidos. Com validation set ocorre uma minimização do erro quadrático mínimo face à época em questão produzindo um modelo genérico consistente.

2.5)

Sem Validação:



Com Validação:



Neste caso, aumentamos o número de unidades na camada escondida. É possível concluir do output que o modelo não nos consegue dar uma previsão geral com novos dados porque está restringido pelos dados anteriores, temos então overfitting neste modelo.

Com validação obtemos sempre uma melhoria nos dados como é possível verificar no exemplo acima.

Se o número de unidades na camada for muito baixo não conseguimos obter uma boa previsão uma vez não tem capacidade para dar uma boa previsão, necessita de mais nós.

É possível concluir que existe um número ótimo de nós de forma a obter um modelo genérico consistente para o problema em questão.