## Chapter 6

# Simulations and Results

From now on, the five configurations for ANN found in the chapter before, will be discussed. They will be simulated for different sets of data, their training evolution compared (avoiding overfitting, and local minimuns) to find the most appropriated number of epochs, and their weights/biases distribution compared, trying to find a network which always tends to evolve to the same parameters, because this will be regarded as a good generalizator.

In the second section of this chapter, GAs will again be implemented, but this time to train the networks from a random state to (what at least is expected to be) the optimal solution.

#### 6.1 Training with learning algorithms

The five configurations were trained with 2000 epochs, because this number is thought to be high enough, and after that with matlab *plotperf* function, the performance evolution was represented. In Figures 6.1 and 6.2 the output of this function for the first and the second configuration are represented and in Appendix B for the other three configurations after training. One can observe that some of them do not train all the epochs, this is because the early stopping criterion, 100 validation failures.

Before the observation of these performance figures, it is important to mention

that matlab, when a network is trained with a set of data, makes three partitions of this set, choosing the first for training (60%), the second for validating (30%) and the third for testing (10%), and the performance of all them is represented in the figures.



Figure 6.1: Performance evolution of configuration 1. 4000 Epochs.

After training the five configurations, they were simulated for a set of 4000 patterns (4000 wrap. csv), and the output error of the estimations is represented here in the next figures and the MAE and STD of error computed.

Analyzing the results (MAE and STD), all the configurations are almost equally as good in deriving estimates, so it the performance evolution will be looked at in order to find the best of these configurations.

Looking at the performance figures, all the configurations tends to be approximately the same performance value, but if it is observed the Figures 6.1, 6.2 and B.7, they all converge faster than the rest. This is due to the used algorithm in this case, *Resilient Backpropagation*. It can be concluded that an adjustment of the learning rate during the training periode, improve considerably the performance of an ANN.



Figure 6.2: Performance evolution of configuration 2. 4000 Epochs.



Figure 6.3: Output error for 4000 patterns with configuration 1. MAE = 2,8125and STD = 3,4445



Figure 6.4: Output error for 4000 patterns with configuration 2. MAE = 2,9058 and STD = 3,4970



Figure 6.5: Output error for 4000 patterns with configuration 3. MAE = 3,0163and STD = 3,6649



Figure 6.6: Output error for 4000 patterns with configuration 4. MAE = 3,0185and STD = 3,4451



Figure 6.7: Output error for 4000 patterns with configuration 5. MAE = 2,9512and STD = 3,4786

## 6.2 Training with genetic algorithms

Another genetic algorithm was designed, but this was not to determine the best configuration of a Neural Network, but to train it. This time the chromosomes were something different than in Section 5.2.

Each individual represents a network, being the chromosomes a distribution of weight and bias values for the five selected configurations of networks. Maybe the fact that any attention is being paid to the learning algorithm previously selected, sounds a little strange, but if the supposition that is not so relevant to the learning algorithm as the rest of parameters in the configuration is made, we can go on with these configurations.

The implemented algorithm will have the same structure as in 4. Three functions in Matlab were defined in the same way to Section 5.2: *Main Genetic.m*, *Reproduction.m* and *Evaluation.m*, which can again be found in the enclosed documents.

The supposition that weights and bias were not bigger than abs(1,5) was taken for the simulations. The selected parameters for the simulations are presented below:

- Number of individuals in population: 20
- Number of children in each reproduction: 7 (It is odd because in each reproduction the last children will be created by a different method. Not reproduction, but averaging all the individuals in the population)
- Number of generations: 1000
- Early stop criterion: 20 generations with no improvements
- Number of individuals in selection procedures (tournaments): 3
- Probability of cross-over: 80%
- Probability of mutation: 50%
- Fitness function: 0, 6(STD) + 0, 4(MAE)

Each configuration, has a different number of weight and bias values, depending on the number of hidden units and on the presence or absence of shortcut connections. The individuals in the first configuration have 74 genes, in the second 134, in the third 211, in the fourth 76 and in fifth 301. So now, if the possesion of optimal results is wanted for all them, it is needed that the reproduction functions affect in the same way to all them. This is because if there is a mutation (*Probability* = 0, 5), it is important that the number of mutated chromosomes for each configuration, will be dependent on the number of chromosomes in the individuals. Half of the chromosomes multiplied by a random number belonging (0, 1) will be mutated. The randomness is always introduced in this algorithms, to preserve the evolutive theories.

From now on the results for all the configurations will be presented, but first it is interesting to mention that although these simulations take a long time to converge (almost two hours for 1000 generations), they are expected to be really good generalizators, due to the fact that they are learning randomly from the fitness variable, and not for each error value but in a more general way.

In Figure 6.8 the MAE and STD evolution for the first configuration are presented, for 1000 generations, with a set of 2000 patterns for simulations (*wrap2000.csv*). The rest of the Figures are included in Appendix B. The convergence of the Algorithm can be observed.



Figure 6.8: MAE and STD evolution with GAs

After training all the configurations, it was observed that all the individuals in each of the populations were very similar, or almost the same, this is because they all have converged to an optimal distribution of parameters (weights and biases). These distributions of parameters are presented in the enclosed documents.

Then each population was averaged, to obtain five parameter distributions and

after, simulation of them with another set of patterns, this time wrap4000.csv, and here in Figure 6.9 represents the output error for the first configuration and in Appendix B the rest.



Figure 6.9: Output error for 4000 patterns with GAs. Configuration 1.  $MAE = 3,0866 \ STD = 4,1451$ 

## 6.3 Conclusions of this Chapter

Although the networks trained with genetic algorithms were expected to be the best estimators, after the simulations, it was observed that they were good enough, but not better than the networks trained with the previously presented (Section 3.3) algorithms. Maybe the fact that these configurations were optimal for a particular learning algorithm leave (in this case) the genetic algorithms in the background, or may be the reproduction functions of the GAs (mutation, cross over) should be modified (probabilities, number of maximun mutations, cross points,...), trying to find an optimal solution. Anyway it is interesting to observe, how since a random network, without backpropagation of error, and complex mathematical equations a network can be trained attending only to a fitness value. This fitness value leaves the researchers a lot of possibilities to train networks attending to many different performance variables, inclusive combining some variables, or introducing penalizations (even for trying to minimize the number of input parameters).

From now on the results obtained with the five selected configurations will be compared with the standards of presents blood pressure devices. These standards are for sphygmomanometers devices, but they can give us an idea of how good are the ANN estimators.

In 1987, the American Association for the Advancement of Medical Instrumentation (AAMI), published a standard for sphygmomanometers, which included a protocol for evaluating the accuracy of devices. In 1990, a protocol was devised by the British Hypertension Society, *BHS*, and later by the European Society of Hypertension *ESH*. This was then used as the basis for the International Protocol [30].

Attending to the British protocol the MAE between different measurements (ANN estimations and IBP in our case) represent the accuracy of the device. It falls into four grades, A to D, where A denotes the greatest and D the least agreement. These grades represent the cumulative percentage of readings falling within 5, 10 and 15 mmHg. To fulfil the BHS protocol, a device must achieve at least grade B. Grades are specified separately for DBP and SBP [5]. However, as it has been only designed the ANN for estimating SBP that is the variable which is here of interest. The different grades are represented in Table 6.1.

Grade	A	В	С	D
$<5 \mathrm{~mmHg}$ (%)	60	50	40	Worse than Grade C
<10 mmHg (%)	85	75	65	
<15  mmHg  (%)	95	90	85	

Table 6.1: Grading criteria used by the BHS

According to the results represented in Table 6.2, these five configurations are expected to be very good estimators of SBP (all tests passed with grade A). Futures researching directions must be focused in trying to find estimators for the other BP parameters (DBP, MBP and PP).

Put together by a working group of the European Society of Hypertension, the

Configuration	$<$ 5 mmHg (%)	< 10 mmHg (%)	< 15 mmHg (%)
1	85,88	98,28	99,7
1 with GAs	82,27	96,33	99,25
2	83,98	98	99,68
2 with GAs	74,55	$96,\!95$	99,18
3	82,75	98	99,58
3 with GAs	78,85	97,2	99,32
4	82,27	97,68	$99,\!65$
4 with GAs	77,88	97,88	99,55
5	82,85	97,95	99,62
5 with GAs	70,97	94,4	99,12

Table 6.2: Evaluation of the configurations with the BHS standard

International Protocol for Validation of Blood Pressure Measuring Devices in Adults provides a recommendation for the validation of BP measuring devices. These accuracy criteria are quite similar to those of the BHS. Categories for rounded values are: 5 mmHg (no error of clinical relevance), 6-10 mmHg (slightly inaccurate), 11.15 mmHg (moderately inaccurate) and 15 mmHg (very inaccurate).

According to this last criteria the selected ANN have in the majority of estimations, no error of clinical relevance. It also true that the BHS standard is not exactly evaluated in this way, but along different periods of time, and with different patients. So next tasks would be to test this ANN direct with patients and check if there are still good estimators, and the most important if they behave properly in different situations.