# Optimization in terms of H1 and H2

**parameters :**
size of validation set : 1/3 of training set, **1800 tupples**
leerning rate : **constant, 0.01**
momentum : **0.1**
early stopping : done when the error on validation set, averaged on 5 epochs, stops decreasing significantly (significanty : new values is above 4/5 of the old value)
number of redondancy **: 10** (for each combination of H1 and H2 we launch the computation of the error 10 times, this is why the following values are averaged)

error on validation set at the end of the process (averaged over 10 redondancy)

| H1 \ H2 | 2 | 3 | 5 | 8 | 10 | 12 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.454 | | | | | | | |
| 5 | 0.420 | 0.230 | 0.041 | | | | | |
| 8 | 0.417 | 0.224 | 0.015 | 0.020 | | | | |
| 10 | 0.418 | 0.222 | 0.019 | 0.014 | 0.023 | | | |
| 12 | 0.419 | 0.224 | 0.015 | 0.015 | 0.015 | 0.013 | | |
| 15 | 0.418 | 0.225 | 0.016 | 0.013 | 0.013 | 0.012 | | |
| 20 | 0.415 | 0.219 | 0.013 | 0.013 | 0.013 | 0.012 | 0.013 | |
| 25 | 0.415 | 0.216 | 0.014 | 0.014 | 0.014 | 0.012 | 0.014 | |
| 50 | 0.433 | 0.224 | 0.012 | 0.012 | 0.011 | 0.011 | 0.011 | 0.011 |
| 75 | NaN | NaN | 0.012 | 0.011 | 0.011 | 0.010 | 0.010 | 0.010 |

standard deviation on error on validation set at the end of the process

| H1 \ H2 | 2 | 3 | 5 | 8 | 10 | 12 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|
| 2 | 0.042 | | | | | | | |
| 5 | 0.012 | 0.015 | 0.018 | | | | | |
| 8 | 0.008 | 0.014 | 0.006 | 0.007 | | | | |
| 10 | 0.009 | 0.007 | 0.009 | 0.004 | 0.024 | | | |
| 12 | 0.009 | 0.007 | 0.004 | 0.006 | 0.005 | 0.002 | | |
| 15 | 0.012 | 0.009 | 0.005 | 0.002 | 0.003 | 0.002 | | |
| 20 | 0.007 | 0.005 | 0.002 | 0.003 | 0.001 | 0.002 | 0.002 | |
| 25 | 0.008 | 0.006 | 0.003 | 0.002 | 0.003 | 0.001 | 0.002 | |
| 50 | 0.063 | 0.033 | 0.002 | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 |
| 75 | NaN | NaN | 0.002 | 0.001 | 0.001 | 0.001 | 0.001 | 0.002 |

number of epochs at the end of the process (averaged over 10 redondancy)

| H1 \ H2 | 2 | 3 | 5 | 8 | 10 | 12 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|
| 2 | 13.0 | | | | | | | |
| 5 | 11.5 | 15.0 | 27.5 | | | | | |
| 8 | 12.0 | 15.0 | 31.5 | 33.5 | | | | |
| 10 | 11.5 | 15.0 | 29.0 | 34.0 | 30.5 | | | |
| 12 | 10.5 | 15.0 | 33.0 | 32.5 | 34.0 | 33.5 | | |
| 15 | 10.0 | 15.0 | 30.0 | 32.0 | 31.5 | 32.0 | | |
| 20 | 11.5 | 15.0 | 33.0 | 31.0 | 29.5 | 31.0 | 32.0 | |
| 25 | 10.5 | 15.5 | 30.0 | 31.0 | 29.0 | 30.5 | 30.5 | |
| 50 | 10.0 | 15.0 | 32.5 | 30.0 | 32.5 | 32.5 | 32.5 | 31.0 |
| 75 | 19.3 | 22.3 | 32.0 | 31.5 | 32.5 | 31.0 | 31.0 | 32.0 |

# Optimization in terms of H1 and H2

Number of misclassified elements of the validation set at the end of the process (indicative)

| H1 \ H2 | 2 | 3 | 5 | 8 | 10 | 12 | 20 | 30 |
|---|---|---|---|---|---|---|---|---|
| 2 | 689.3 | | | | | | | |
| 5 | 672.5 | 367.0 | 39.8 | | | | | |
| 8 | 663.5 | 372.7 | 9.4 | 13.0 | | | | |
| 10 | 699.9 | 360.1 | 8.7 | 7.0 | 7.3 | | | |
| 12 | 681.5 | 379.1 | 6.9 | 5.6 | 5.8 | 3.9 | | |
| 15 | 654.7 | 338.9 | 6.5 | 3.6 | 3.0 | 3.2 | | |
| 20 | 725.5 | 317.6 | 3.4 | 2.3 | 2.6 | 2.0 | 2.2 | |
| 25 | 700.2 | 367.1 | 2.2 | 1.7 | 2.4 | 0.8 | 1.6 | |
| 50 | 715.0 | 373.5 | 0.8 | 0.2 | 0.1 | 0.2 | 1.0 | 0.0 |
| 75 | 949.2 | 581.2 | 0.2 | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 |

Note :
For H1 = 75 there are two values for H2 that have caused the gradient descent not to converge reducing the learning rate at 0.001 makes it converge each time, we can them conclude that it is a "normal" divergence due to a too high learning rate in this configuration of parameters.