OMML homework 1 report

jad zouwayhed , Giorgio Zannini Quirini

November 25, 2019

1 Question 1

1.1

- parameters
  - $N = 128$
  - $\sigma = 6.5$
  - $\rho = 10^{-5}$.

  Parameters were chosen via grid search on the validation set. Too high values of N would cause overfitting because neurons would explain just a few or a single observation losing their generalization capacity, viceversa for underfitting.
- scipy toolbox parameters
  - method = BFGS (tolerance = 10^{-5}, max iterations = None)
  - successful optimization
  - number of iterations = 2075
  - number of function evaluations = 2160
  - number of gradient evaluations = 2160
  - starting value of function = 43.2878
  - final value = 0.00032
  - gradient norm at the start = 102.4126
  - gradient norm at the end = 4.0319×10^{-5}
- train errors
  - initial value of train error = 60.3054
  - final value of train error = $1.6955567370624064 \times 10^{-5}$
- other errors

  - validation error = $3.1450974808630916 \times 10^{-5}$
  - test error = $7.322666110383915 \times 10^{-5}$

- comparison
  - In comparison to the Rbf network, Mlp performs much better in terms of errors. On the other hand, it's slower since it requires many more function and gradient evaluation to reach its goal.

1.2

- parameters
  - $N = 35$
  - $\sigma = 0.55$
  - $\rho = 10^{-5}$
- scipy toolbox parameters
  - method = BFGS (tolerance = 10^{-5}, maxiterations = None)
  - successful optimization
  - number of iterations = 852
  - number of function evaluations = 861
  - number of gradient evaluations = 861
  - starting value of function = 1.2729
  - final value = 0.004031841917793968
  - gradient norm at the start = 3.7913
  - gradient norm at the end = $3.429568487630391 \times 10^{-5}$
- train errors
  - initial value of train error = 1.3383
  - final value of train error = 0.001104814134504591
- other errors
  - validation error = 0.005821368184085043
  - test error = 0.010021050780969317

2 Question 2

2.1

- Parameters
  - $N = 128$
  - $\sigma = 6.5$
  - $\rho = 10^{-5}$
  - method = BFGS (tolerance = 10^{-5}, maxiterations = None)
  - seed = 884633801
- Errors
  - training error : 0.04915940897294787
  - test error : 0.07684315652396305
- In comparison to MLP in Question 1, Both the train and the test error in Question 1 are much smaller being of order 10^{-5}, hence, in terms of results, classical MLP is significantly better.
  However, in terms of effeciency, ELM outperforms MLP significantly as MLP on average takes 16 to 20 seconds while ELM takes less than one second.

2.2

- Parameters
  - $N = 35$
  - $\sigma = 0.55$
  - $\rho = 10^{-5}$
  - method = BFGS (tolerance = 10^{-5}, maxiterations = None)
  - seed = 277939856
- Errors
  - training error : 0.0930732779267787
  - test error : 0.04944513298317685

3 Question 3

- parameters:
  the tolerances ξ_1^k and ξ_2^k are updated with θ at every iteration k
  - $N = 128$
  - $method_1$ = BFGS
  - $method_2$ = BFGS
  - $\sigma = 6.5$
  - $\rho = 10^{-5}$
  - $\xi_1^0 = 10^{-7}$
  - $\xi_2^0 = 10^{-6}$
  - $\theta = 0.6$
  - maxiter(second block "W b") = 5000
- Stopping criteria:
  The algorithms stops when the norm of the gradient of the second block (non convex) is below a threshold τ
  - $\tau = 10^{-9}$
- Exeution Details:
  - Number of outer iterations = 16
  - Number of function evaluations = 24589
  - Number of gradient evaluations = 24589
  - Computational time = 298.2 seconds
- Errors:
  - Train Error = $7.171732266551507 \times 10^{-6}$
  - Test Error = $2.8334444443316687 \times 10^{-5}$
- Comparison with extreme learning
  - In terms of quality, two-block decomposition has a very low test error compared to extreme learning in question 2 by 3 orders of magnitude. However, in terms of efficiency, extreme learning is faster after obtaining the seed for optimal initial values.

EX	FFN	settings	train error	val error	test error	Opt time
Q1.1	Full MLP	N = 128,$\sigma = 6.5$, $\rho = 10^{-5}$	1.695e^{-5}	$3.145e^{-5}$	$7.322e^{-5}$	41.6s
Q1.2	Full RBF	N = 35,$\sigma = 0.55$, $\rho = 10^{-5}$	0.0011	0.0058	0.0100	0.49s
Q2.1	Extreme MLP	N = 128,$\sigma = 0.55$, $\rho = 10^{-5}$	0.124	0.0865	0.0494	77s
Q2.2	Extreme RBF	N = 35,$\sigma = 0.55$, $\rho = 10^{-5}$	0.09307	0.0374	0.005	69s
Q3	Two-Block MLP	N = 128,$\sigma = 6.5$, $\rho = 10^{-5}$	$7.171e^{-6}$	$1.271e^{-5}$	$2.833e^{-5}$	298s

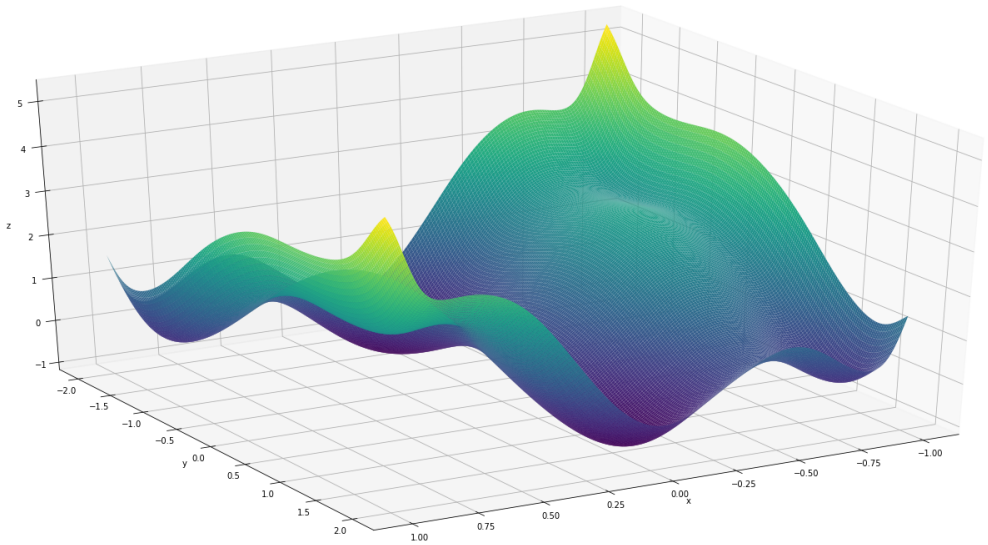

Figure 1: FULL MLP

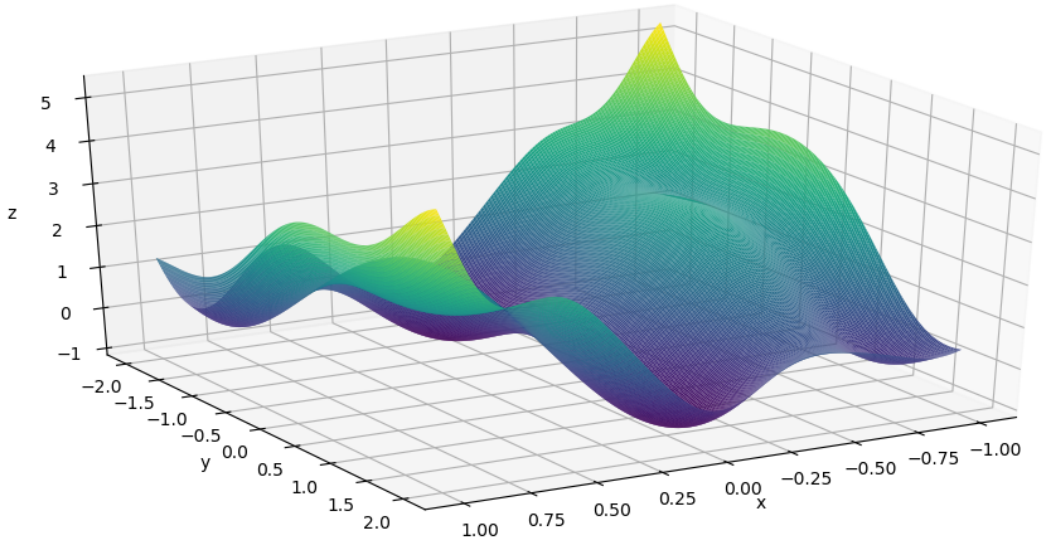

Figure 2: FULL RBF

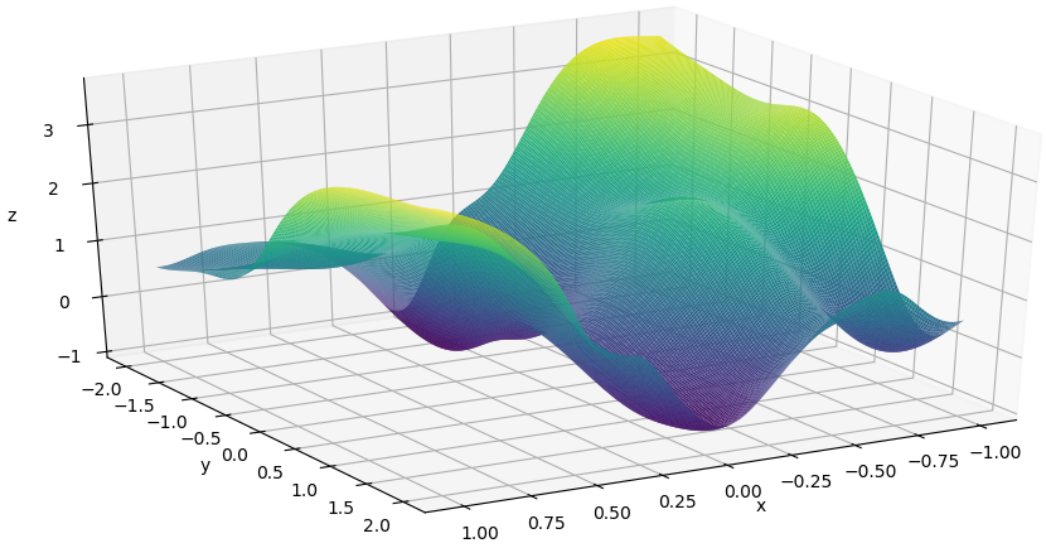

Figure 3: Extreme Learning MLP

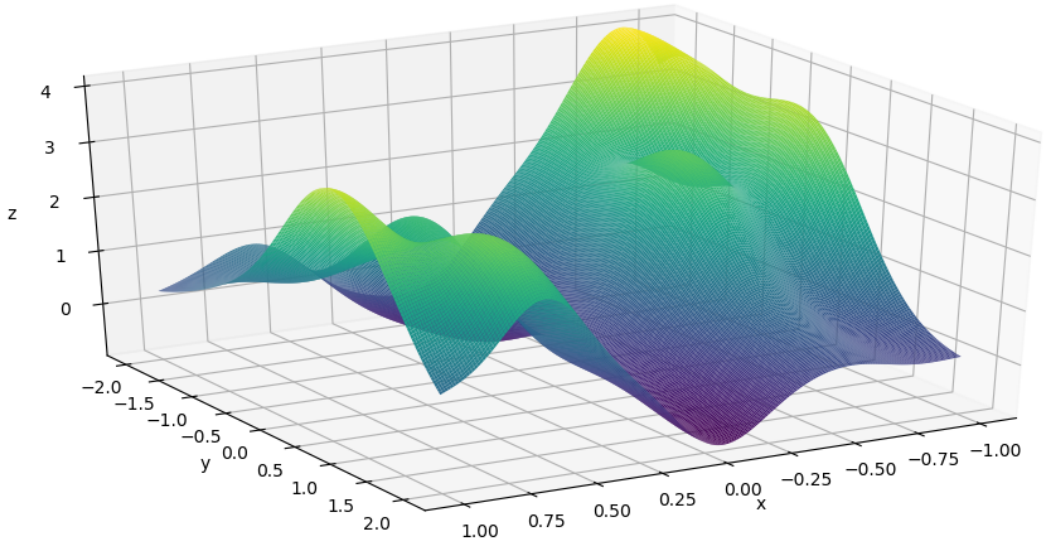

Figure 4: Unsupervised Centers RBF

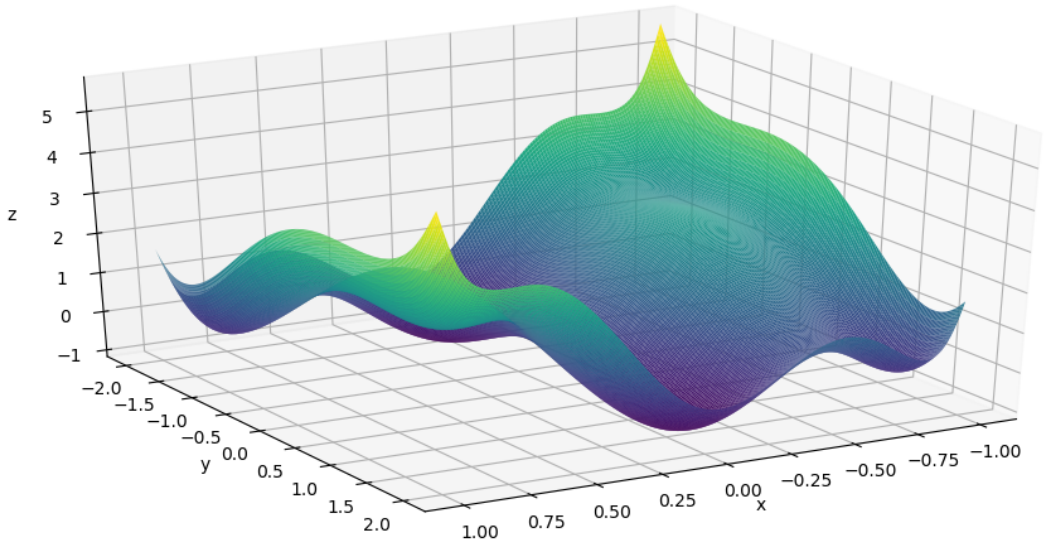

Figure 5: Two-block decomposition MLP