

# **OMML Assignment**

## **MLP and Generalized RBF Network**

**Submitted by:**  
**Sibghat Ullah**  
**Maryam Zamanipour**  
**Siddhant Tandon**

# OMML REPORT

## Question 1

### Part 1

#### 1. HyperParameters

**N = 10**

**roh = 0.0001**

**sigma = 0.1**

Over-fitting occurs when Train error is low but test error is high i.e. approximation function found fits the training data better but doesn't fit the test data.

Under-fitting simply means the Train error is high because the Approximating function doesn't fit the train data.

By doing a gridsearch over the hyperparameters, and choosing those hyperparameters over which both train and test errors are the lowest, these hyperparameters become optimal and result in a good fit.

Evidence: Following screenshot shows the train and test errors over values of hyperparameters.

Final loss function values at optimal hyperparameters:

**test\_error': 0.011718666353836201**

**train\_error': 0.009188975569789974**

	19	20	21	22	23	28	29	32	33	34	35	36	37	38	39	40	41	42	45
N	10	10	10	10	10	10	10	10	10	10	10	10	20	20	20	20	20	20	20
roh	0.0001	0.0001	0.0001	0.01	0.01	0.1	0.1	3	3	0.0003	0.0003	0.0003	0.0001	0.0001	0.0001	0.01	0.01	0.01	0.05
sigma	0.001	0.1	0.05	0.001	0.1	0.001	0.1	0.1	0.05	0.001	0.1	0.05	0.001	0.1	0.05	0.001	0.1	0.05	0.05
test_error	0.024351934	0.0117186664	0.0181785173	0.1532745876	0.0838142	0.1532745833	0.1532746099	0.1532816143	0.1532757115	0.0384856793	0.0193460181	0.0202336355	0.0242033127	0.0178717827	0.0181955782	0.1532745875	0.083759456	0.1064789485	0.1532745926
train_error	0.0214887352	0.0091889756	0.0153335213	0.1587525796	0.0814931133	0.1587525855	0.158752602	0.1587596065	0.1587537036	0.0355679124	0.0164922357	0.0173762322	0.0215156798	0.015033182	0.0153441644	0.1587525796	0.081435943	0.1054513211	0.1587525847

#### 2. Optimization routine

Method: BFGS

max-iterations: 30,000

Training objective function: 0.01444

Test MSE: 0.01557

Training computing time: 11.23 seconds

Function evaluations: 36834

Gradient evaluations: 877

#### 3. Training and Test errors

Training objective function: 0.01444

Test MSE: 0.01557

## Part 2

### 1. HyperParameters

**N = 20**

**roh = 0.0001**

**sigma = 0.1**

Over-fitting occurs when Train error is low but test error is high i.e. approximation function found fits the training data better but doesn't fit the test data.

Under-fitting simply means the Train error is high because the Approximating function doesn't fit the train data.

By doing a gridsearch over the hyperparameters, and choosing those hyperparameters over which both train and test errors are the lowest, these hyperparameters become optimal and result in a good fit.

Evidence: Following screenshot shows the train and test errors over values of hyperparameters.

	19	20	23	24	25	26	27	28	37	38	39	40	41	42	43	44	45	46	47	48	49
N	10	10	10	10	10	10	10	10	20	20	20	20	20	20	20	20	20	20	20	20	20
roh	0.0001	0.0001	0.01	0.01	0.05	0.05	0.1	0.0001	0.0001	0.0001	0.01	0.01	0.01	0.05	0.05	0.05	0.1	0.1	0.1	0.1	3
sigma	0.001	0.1	0.1	0.05	0.001	0.1	0.05	0.001	0.001	0.1	0.05	0.001	0.1	0.05	0.001	0.1	0.05	0.001	0.1	0.05	0.001
test_error	0.1541348633	0.0572771937	0.0826665199	0.1518960721	0.2712704393	0.15094972	0.2019735715	0.3069388056	0.1537476006	0.037945496	0.1886733051	0.1966521714	0.0721175723	0.1395375014	0.3493078164	0.1739616305	0.2913193179	0.5652194322	0.2963063168	0.4183771478	2.3353999644
train_error	0.1454742402	0.0201109648	0.0537326207	0.1223626414	0.2649579444	0.1274603319	0.1720457723	0.3068356682	0.1342989923	0.0036347449	0.0437533844	0.1779885006	0.0445995374	0.1018792104	0.3317338533	0.1432632276	0.2423204353	0.5480040811	0.2690089502	0.3578649474	2.3408779566

Final loss function values at optimal hyperparameters:

**test\_error': 0.029392371360539207**

**train\_error': 0.0046727970828470515**

### 2. Optimization routine

Method: Nelder-Mead

max-iterations: 30,000

Training objective function:0.0088677

Test MSE:0.057829

Training computing time:50.20

Function evaluations:33698

Gradient evaluations: 0 ( Nelder-Mead doesn't use gradient )

### 3. Training and Test errors

Training objective function: 0.0088677

Test MSE: 0.057829

## Question 2

### Part 1

#### 1. HyperParameters

**N = 20**

**roh = 0.0001**

**sigma = 0.1**

Final loss function values at optimal hyperparameters:

test\_error': 0.017926432537633191

train\_error': 0.015139482893578783

#### 2. Optimization routine

Method: BFGS

max-iterations: 30,000

Training objective function: 0.0292170

Test MSE: 0.034264

Training computing time: 0.1207

Function evaluations: 660

Gradient evaluations: 30

#### 3. Training and Test errors

Training objective function: 0.0292170

Test MSE: 0.034264

## Question 2

### Part 2

#### 1. HyperParameters

**N = 20**

**roh = 0.0001**

**sigma = 0.1**

Final loss function values at optimal hyperparameters:

test\_error': 0.034291308898091712

train\_error': 0.0007897255290297369

#### 2. Optimization routine

Method: BFGS

max-iterations: 30,000

Training objective function: 0.016697192

Test MSE: 0.07177

Training computing time: 0.70179

Function evaluations: 902

Gradient evaluations: 41

#### 3. Training and Test errors

Training objective function: 0.016697192

Test MSE: 0.07177

### Question 3

Network chosen: RBF

#### 1. HyperParameters

**N = 20**

**roh = 0.0001**

**sigma = 0.1**

#### 2. Optimization routine

Method for both blocks: CG

max-iterations: 30,000

Training objective function: 0.001467

Test MSE: 0.071312

Training computing time: 331.19175

Function evaluations: 9046

Gradient evaluations: 4871

Total iterations / sub problems solved : 61

#### 3. Stopping Criteria:

Iterate until Gradient with respect to Centers C falls between the tolerance.

Tolerance = [1e-6, 1e-3]

#### 4. Training and Test errors

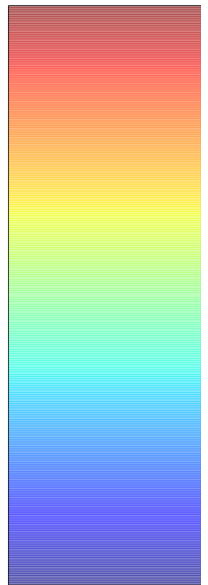
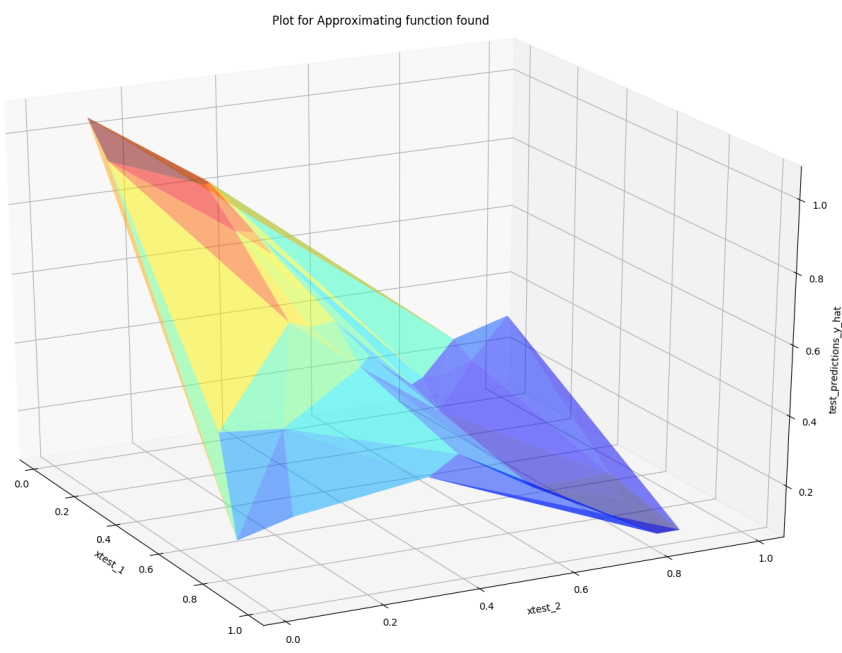
Training objective function: 0.001467

Test MSE: 0.071312

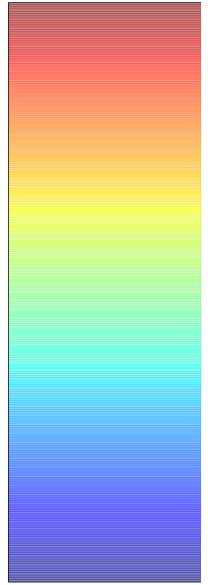
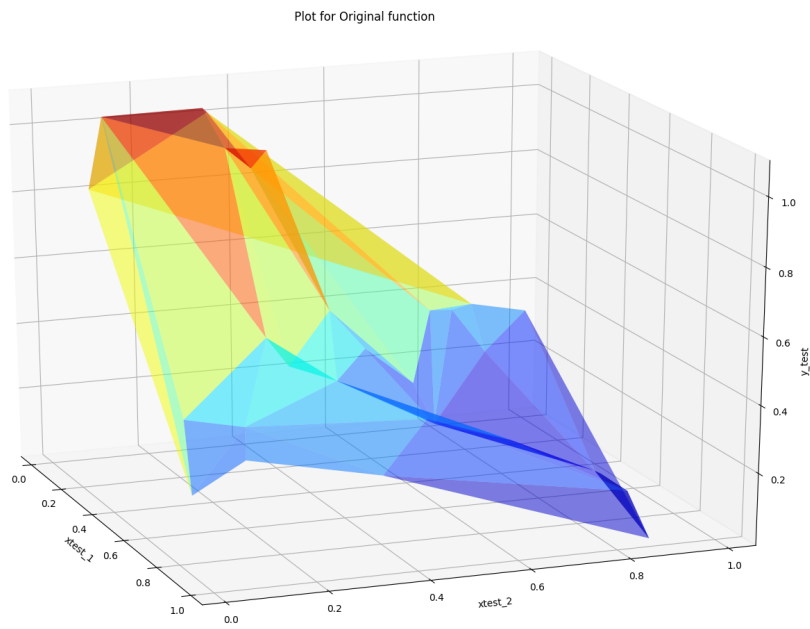
## Final Comparison

Ex	FFN	N	Roh	Sigma	Train error	Test error	Optimization time (seconds)
Q1.1	Full MLP	10	0.0001	0.1	0.01444	0.01557	11.23
Q1.2	Full RBF	20	0.0001	0.1	0.008867	0.057829	50.20
Q2.1	Extreme MLP	20	0.0001	0.1	0.029217	0.034264	0.1207
Q2.2	Unsupervised c RBF	20	0.0001	0.1	0.016697	0.07177	0.70179
Q3	RBF-Two block	20	0.0001	0.1	0.001467	0.071312	331.191

**Q1 part 1**  
Approximating function ( FULL MLP ) :

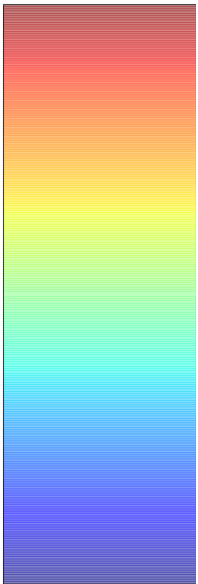
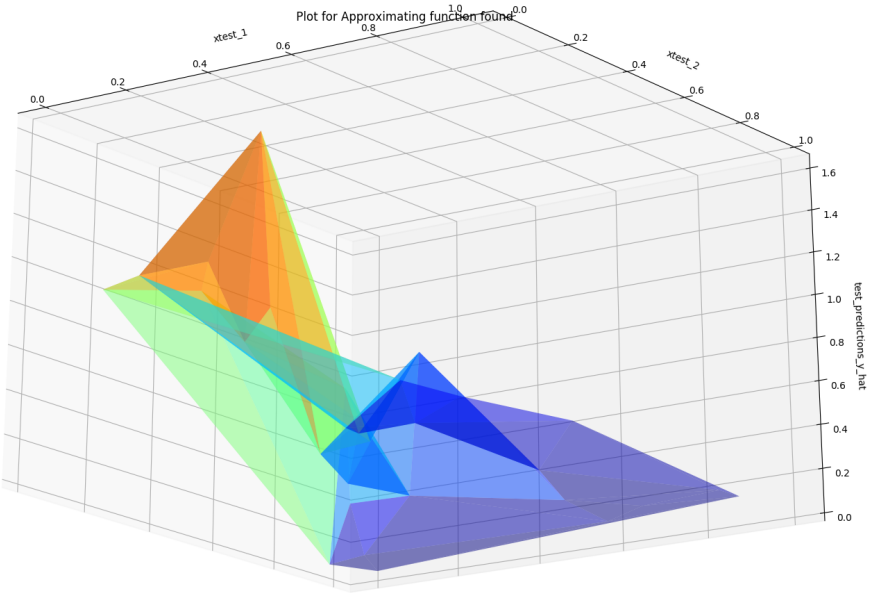


Original function:

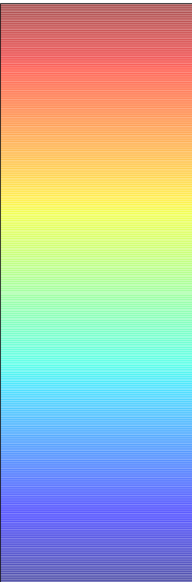
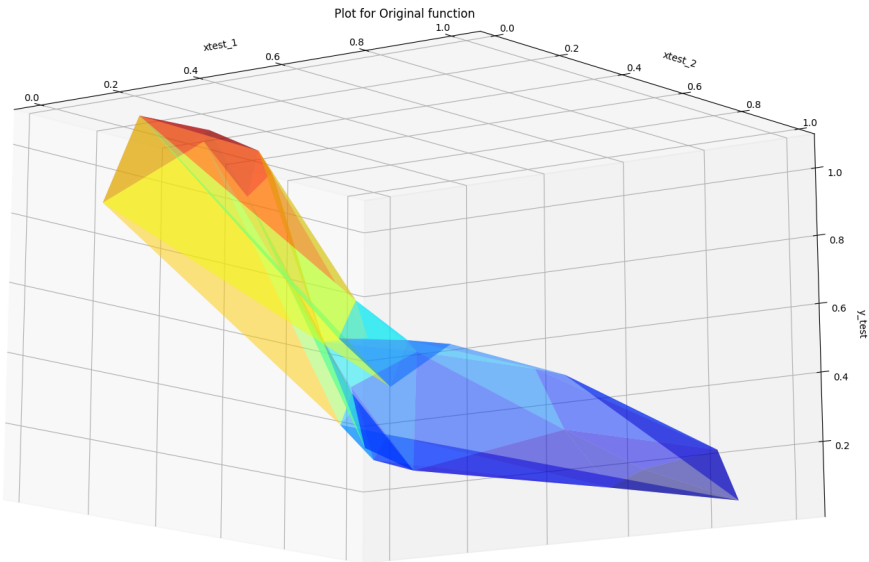




**Q1 part 2**  
Approximating function ( FULL RBF ) :

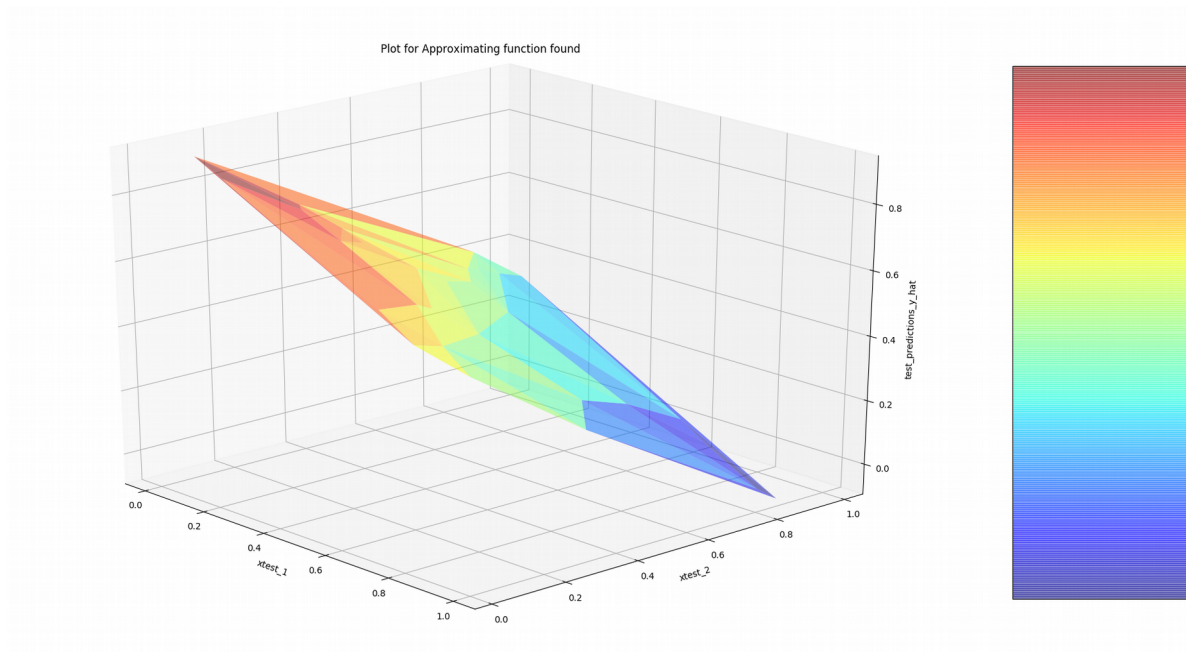


Original function:



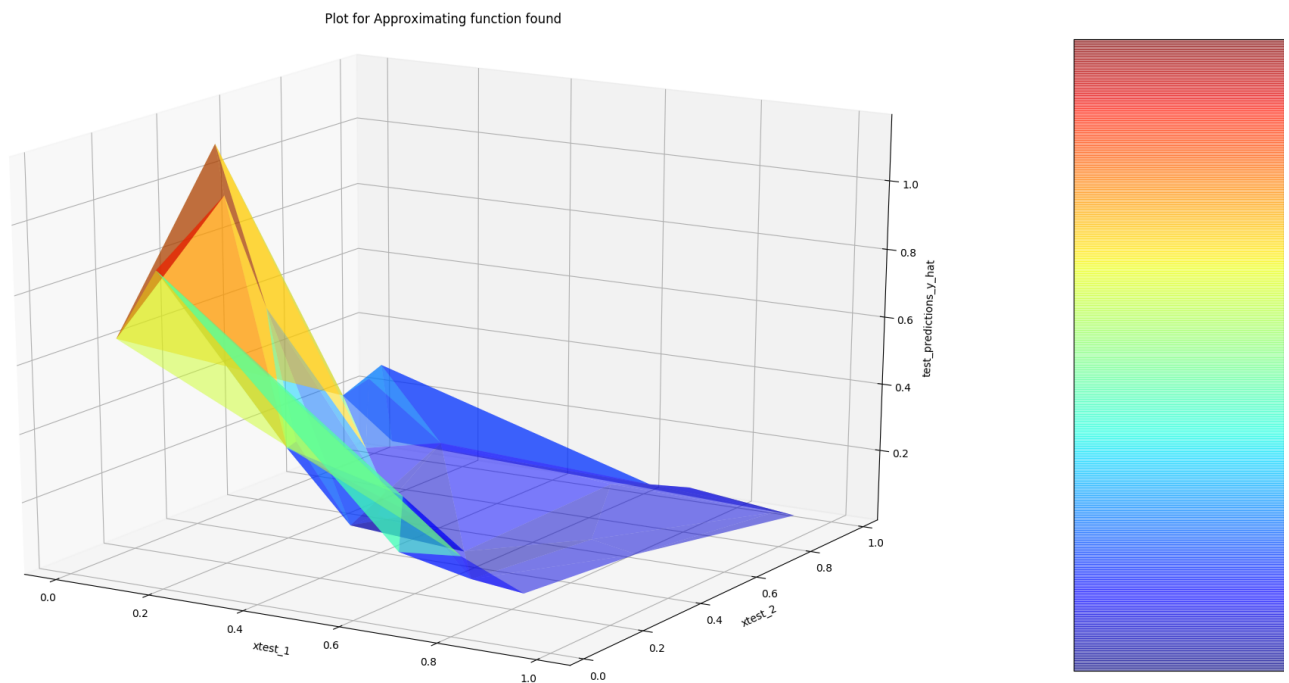
## Q2 part 1

Approximating function ( Extreme MLP ) :

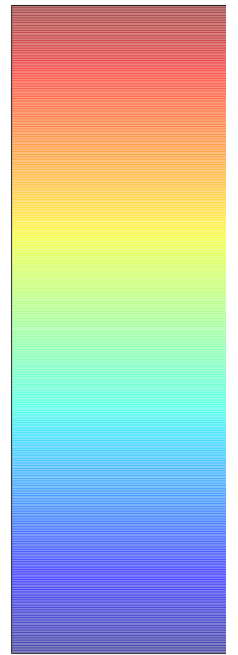
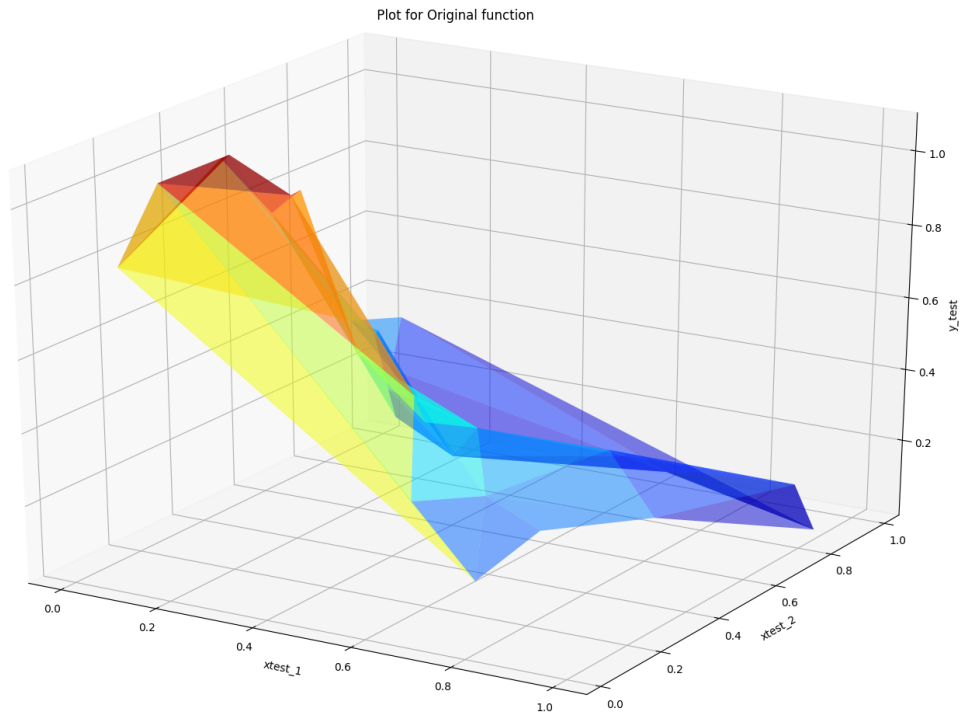


## Q2 part 2

Approximating function ( Unsupervised RBF ) :

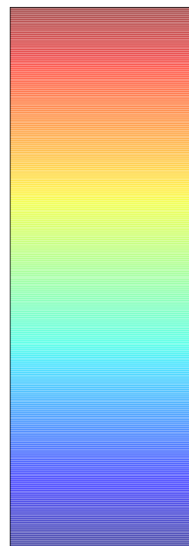
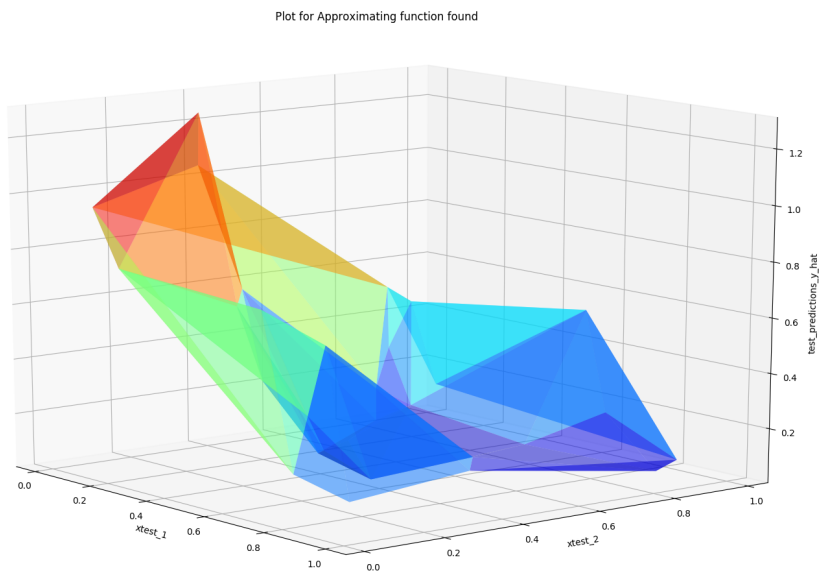


Original function:



**Q3**

Approximating function ( Two blocks RBF ) :



Original function:

