1. What is the lower bound for the number of training examples, N?

- 1 is the min value. We can have any number of samples beyond that.

2. What happens with the error N = n? Why?

- There are as many samples as RBF units/weights. We may find a perfect solution.

3. Under what conditions, if any, does (4) have a solution in this case?

 - If we have n linearly independent equations then we can have solution.

4.During training we use an error measure defined over the training examples. Is it good to use this measure when evaluating the performance of the network?

* No. The error is just used to know how to improve the network. We should use non-training data for evaluating the performance.

-----------------------------------------------------------------------------------------------------------------

Sin(2x)

0.1- 7 units

0.01- 25 units

0.001 – 56 units

Square(2x)

0.1 – 62 units

0.01 - 64 units

0.001 - 64 units

2. The difference between 5 and 6 units is big.

- We need 6 units to capture peaks of sin2x

3.Approximating square(2x) is like Classification.

4. To get a residual value around 0, we need n linearly independent equations.

Number of units = number of x with linear independent equations.

5. Solve the XOR Problem

- Yes we can solve. We can put four units in 4 locations of 00,01,10,11. Each unit triggers based on each of these. Or we can process AND components separately and then feed them to linear separator (single layer perceptron).

1. Residual value = 0.01

unit 60, eta=1, Iter=40000

unit 60, eta=0.25, Iter= 100 000

unit 100, eta=0.35, Iter = 4000

2. exp(x):

unit 100, eta =1, Iter = 800 000, residual error=0.08

unit 500, eta =1, Iter=800 000, residual error = 0.01

Single Winner strategy:

Problem: Sometimes a unit is useless and cover nothing

Advantage: each cluster will get one unit

Single winner strategy:

Some units do not cover anything. Variance is same for all units.

Move all:

All units will cover data. Clusters can have more than one units covering them. Variance of units can be different. Takes time to converge.

Questions

More units you have:

Train set will fit better but test set result are almost the same

Bellow 5 units, there are not enough units to have good result