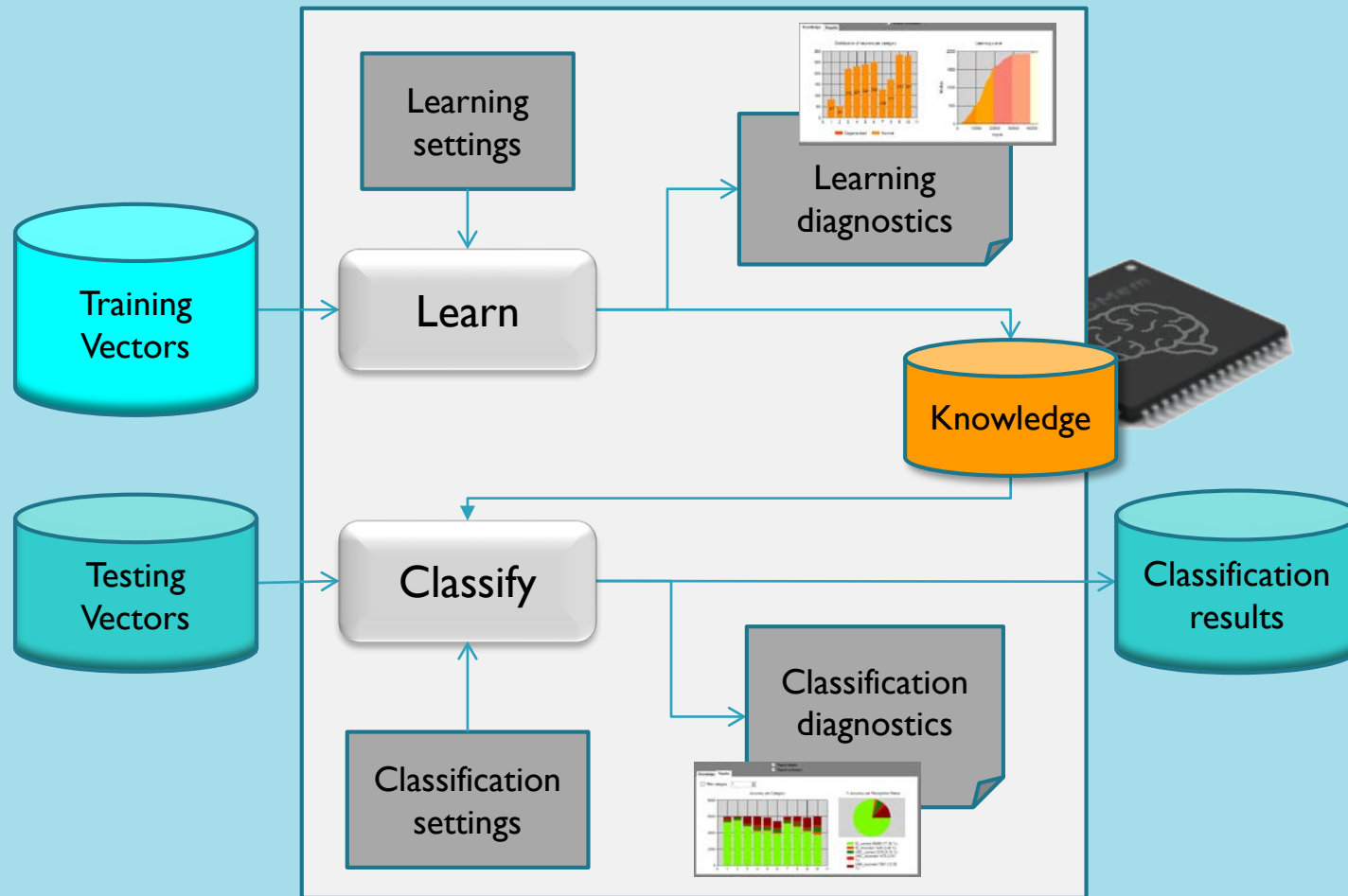


# NeuroMem Knowledge Builder

Learn, Verify, Classify  
...based on a NeuroMem network

# NeuroMem KB diagram



# Objectives for a single feature set (v1)

---

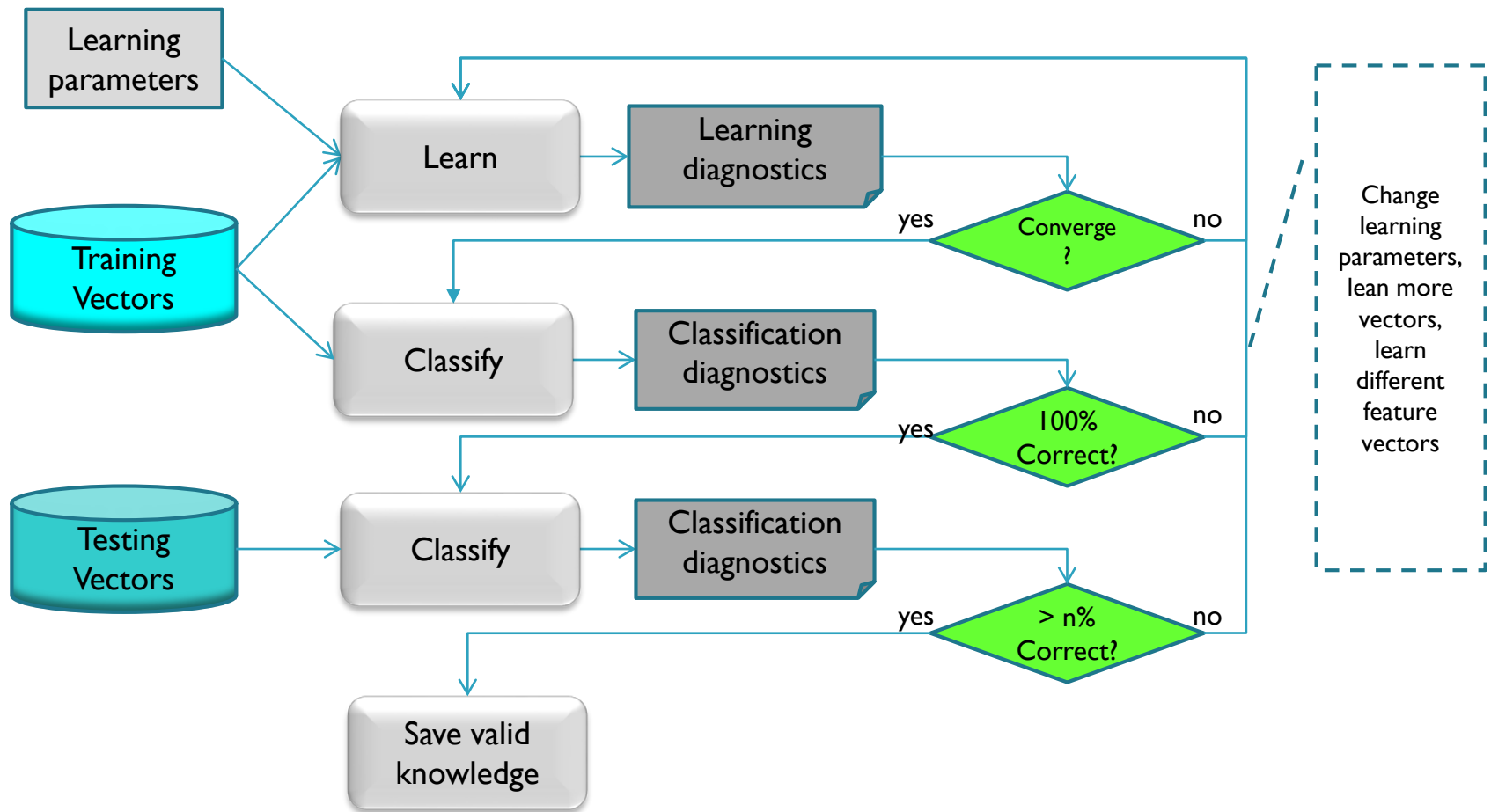
- ▶ Evaluate feature extraction methods
- ▶ Evaluate if a set of training vectors can build a decision space where both generalization and discrimination are satisfactory
- ▶ Identify the categories easy to discriminate, or on the contrary difficult to classify with a single set of feature vectors
- ▶ Identify the pairs of categories introducing confusion
- ▶ Size and validate the accuracy and throughput of a knowledge built on the set of training vectors.
- ▶ Validate the accuracy and throughput of the same knowledge on a set of test vectors.
- ▶ Save and export the knowledge for use on a run-time system

# Objectives for multiple feature sets (v2)

---

- ▶ Rank different feature sets to classify the same objects
  - ▶ ex: a person recognized with its face and voice
  - ▶ ex: an airplane recognized by its overall shape, wingspan and tail
- ▶ Find complementary confusion matrices between feature sets
- ▶ Test decision rules between feature sets
  - ▶ Combinatorial or hierarchical
    - Apply weights, Verify Majority in agreement, Minimum Consensus
- Size and validate the accuracy and throughput of the resulting knowledge
- ▶ Validate the accuracy and throughput of the same knowledge on a set of test vectors
- ▶ Save and export the knowledge for use on a run-time system

# Training and Validation Workflow



# NeuroMem Knowledge Builder version 1.0

Windows based

Single set of feature vectors at a time

Interface to NeuroStack or simulates a chain of CMIK chips

# Main Panel Overview

NeuroMem Knowledge Builder version 1.5.0.0, on SimuNS4K\_4096

File Collect View Context 1 Norm L1 Help

5 of 7

Forget All Learn Classify

Clear Save

Changing NN platform to SimuNS4K\_4096  
NN capacity=4096

File vectors\_tm

Context 1 with Norm L1

| Inputs     | Total | Cat0   | Cat1    | Cat2    | Cat3    |
|------------|-------|--------|---------|---------|---------|
| Percentage | 100 % | 0.00 % | 30.00 % | 18.89 % | 43.89 % |
| Values     | 180   | 0      | 54      | 34      | 79      |

LearnRBF Minif=2, Maxif=16384

| Neurons          | Total | Cat0   | Cat1    | Cat2    | Cat3    |
|------------------|-------|--------|---------|---------|---------|
| Percentage       | 100 % | 0.00 % | 18.18 % | 36.36 % | 27.27 % |
| Values           | 11    | 0      | 2       | 4       | 3       |
| Neurons/Patterns | n/a   | 3.7 %  | 11.76 % | 3.8 %   |         |

Classify using BestMatch category, K=3

| Outputs    | Total | Cat0   | Cat1    | Cat2    | Cat3    |
|------------|-------|--------|---------|---------|---------|
| Percentage | 100 % | 0.00 % | 30.00 % | 18.89 % | 43.89 % |
| Values     | 180   | 0      | 54      | 34      | 79      |

Accuracy

| Correct      | Total | Cat0 | Cat1 | Cat2    | Cat3    |
|--------------|-------|------|------|---------|---------|
| Percentage   | 100 % | 0 %  | 30 % | 18.89 % | 43.89 % |
| > Identified | 100 % | 0 %  | 30 % | 18.89 % | 43.89 % |

File vectors\_tst

Context 1 with Norm L1

| Neurons    | Total | Cat0   | Cat1    | Cat2    | Cat3    |
|------------|-------|--------|---------|---------|---------|
| Percentage | 100 % | 0.00 % | 18.18 % | 36.36 % | 27.27 % |
| Values     | 11    | 0      | 2       | 4       | 3       |

| Inputs     | Total | Cat0   | Cat1    | Cat2    | Cat3    |
|------------|-------|--------|---------|---------|---------|
| Percentage | 100 % | 0.00 % | 34.89 % | 14.93 % | 45.40 % |
| Values     | 3216  | 0      | 1122    | 480     | 1460    |

Learning

Always Forget first

RBF Model Generator  Iterative  
 Load "As Is"  Verify  
 Codebook Generator  
 Clusterize

Min Influence Field: 2  
Max Influence Field: 16384

Classification

K: 3

Category Out

RBF  
 KNN  
 Report confusion

Best match  
 Dominant  
 Unanimity

Distribution per category

| Category | Ground Truth inputs (3216) | Models (11) | Outputs (3216) |
|----------|----------------------------|-------------|----------------|
| 1        | 1122                       | 0           | 1133           |
| 2        | 400                        | 0           | 400            |
| 3        | 1480                       | 0           | 1480           |
| 4        | 154                        | 0           | 154            |

Knowledge Results

Filter category: 1

Accuracy per Category

| Category | ID_correct | UNC_correct | UNC_incorrect | UNK_incorrect |
|----------|------------|-------------|---------------|---------------|
| 1        | ~1100      | ~100        | ~100          | ~100          |
| 2        | ~400       | ~100        | ~100          | ~100          |
| 3        | ~1400      | ~100        | ~100          | ~100          |
| 4        | ~150       | ~100        | ~100          | ~100          |

% Accuracy per Recognition Status

| Status        | Percentage |
|---------------|------------|
| ID_correct    | 87.87 %    |
| UNC_correct   | 11.75 %    |
| UNC_incorrect | 0.34 %     |
| UNK_incorrect | 0.03 %     |

Input: vectors\_tst.csv, Total records: 3216 NN capacity=4096 Learning settings Committed=11 Classif settings 52.14msec

# Data file formats

## Minimum data set

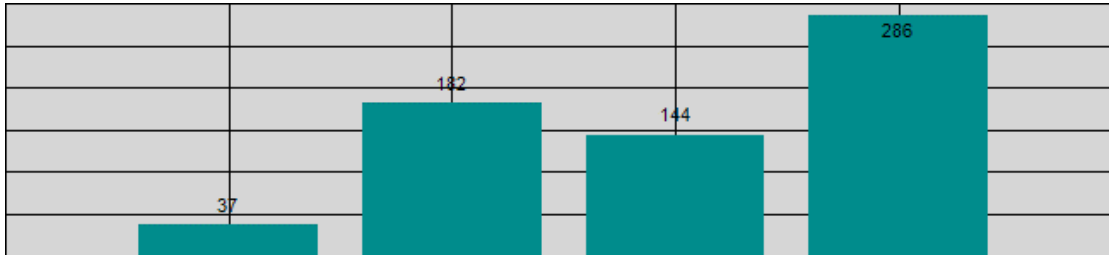
| PatternID                              | ParentID   | Context   | Ground Truth Category  | Vector length L                                       | Vector(0 ... Vector(L))   |
|--|--|---|--|---|---|
| Index of the pattern in the input list | User defined index identifying the origin of the pattern (a file name, XY position, time stamp, etc) | Value identifying the type of feature encrypted in the vector<br><br>Range=[0, 127] | Ground Truth Category associated to the vector<br>0xFFFF otherwise<br><br>Range=[1,327 68] | Number components in the vector<br><br>Range=[1,25 6] | L values of the vector components<br><br>Values ranging between [0,255] |

## Expanded dataset after classification

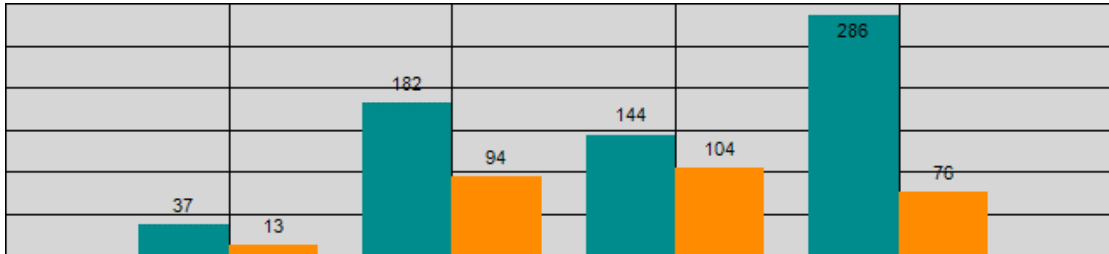
| Status code | Accuracy code | CatOut   | Cat 1  | Dist 1 | Nid1 | Cat 2                             | Dist 2 | Nid2 | ... | CatK  | DistK | NidK |
|-------------|---------------|----------|--|--------|------|-----------------------------------|--------|------|-----|---|-------|------|
| UNKnown     | Correct       | Global   | Response of the 1st firing neuron including its distance, category and identifier. |        |      | Response of the 2nd firing neuron |        |      |     | Response of the K <sup>th</sup> firing neuron |       |      |
| IDentified  | Incorrect     | response |  |        |      |                                   |        |      |     |   |       |      |
| UNCertain   | N/A           | e        |  |        |      |                                   |        |      |     |   |       |      |

In KNN mode, there are always K firing neurons  
In RBF mode, K is a maximum applicable

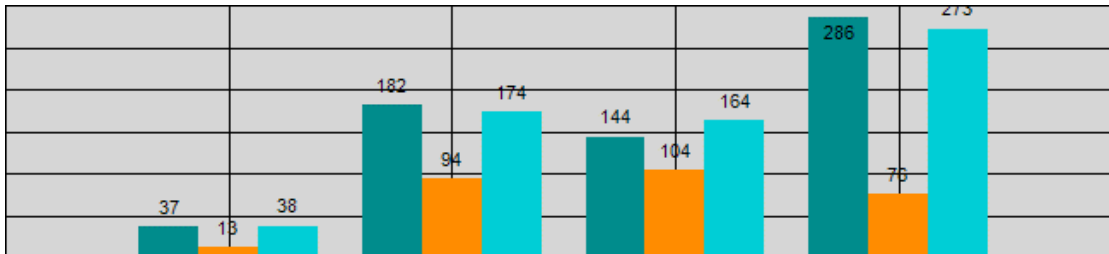




Inputs:  
 Vectors, with associated  
 Ground Truth categories



Models:  
 Inputs retained by the  
 neurons during learning.  
 They constitute a  
 knowledge

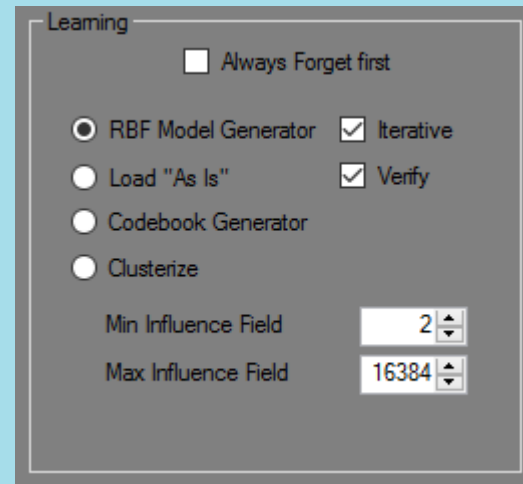


Outputs:  
 Vectors and their  
 categories recognized by  
 the neurons

# Learning Settings

---

- ▶ Methods
- ▶ Parameters
- ▶ Options



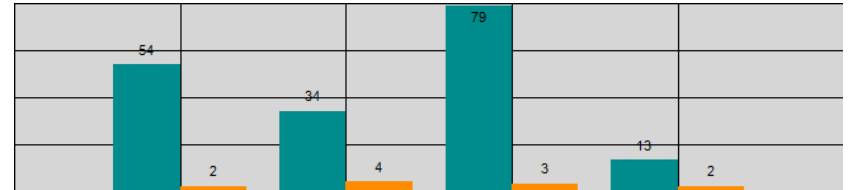
# Learning functions

---

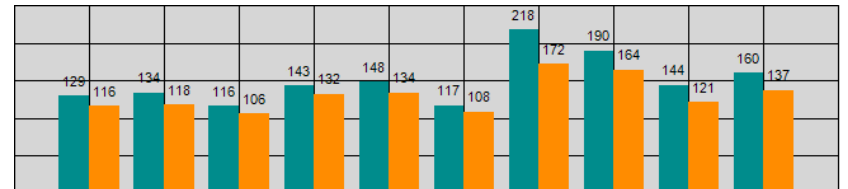
- ▶ **Learning operations**
  - ▶ Learning vectors with their Ground Truth categories
  - ▶ Learning with automatic category allocation (codebooks)
- ▶ **Learning options**
  - ▶ Iterative function by default
  - ▶ Conservative, Moderate, or Liberal (Maximum Influence Field)
  - ▶ Level of Sensitivity (Minimum Influence Field)
  - ▶ Selection of a distance Norm
- ▶ **Learning Diagnostics**
  - ▶ Distribution of the neurons per category and per context
  - ▶ Distribution of the degenerated neurons
  - ▶ Learning curve convergence

# Inputs versus Models ratio

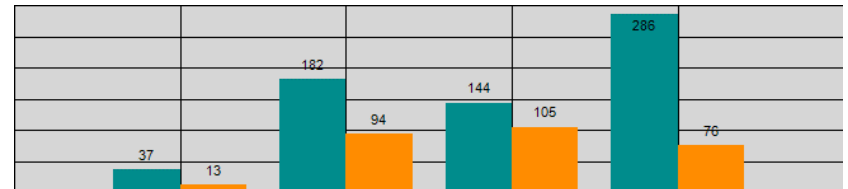
Good generalization across all categories:  
A few neurons model a large number of inputs.  
What accuracy will they deliver on a testing set?



Poor generalization across all categories:  
A ratio (Models ÷ Input) too close to 1.  
Does the ratio improves when learning more  
Inputs?



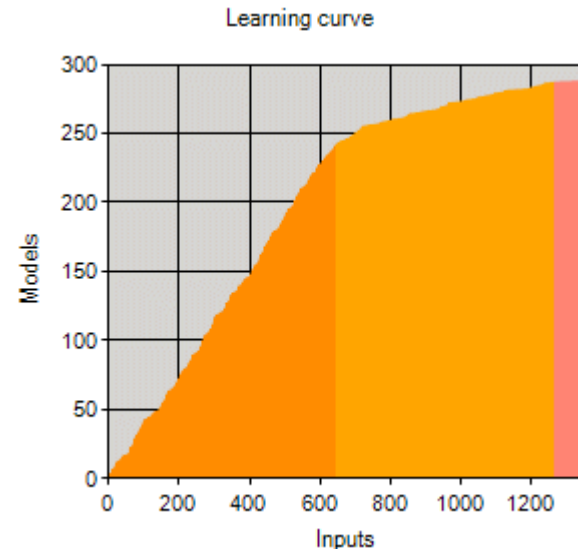
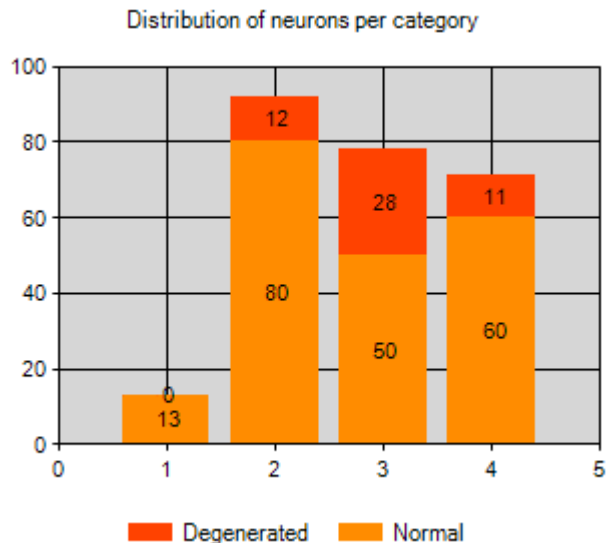
Uneven generalization across categories:  
Accuracy and confusion may vary depending on  
the categories



# Knowledge profile

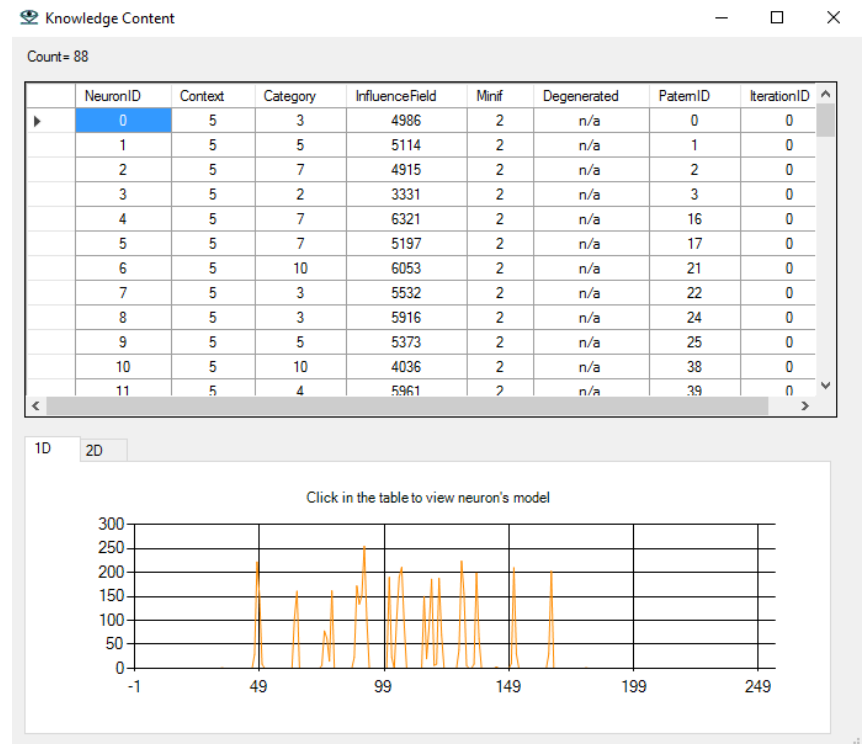
- ▶ Neurons per categories
- ▶ Neurons degenerated
  - ▶ Warning that the boundaries of some categories are overlapping

- ▶ Learning curve should be asymptotic
  - ▶ If not be suspicious of the following...
  - ▶ The network is full and learning could not be completed
  - ▶ Can some GT categories be erroneous ? (this usually also creates degenerated neurons)
  - ▶ Can the feature vectors be inappropriate for the intended classification?
- ▶ Learning is iterative by default to build a decision space invariant to the order of the learned vectors (one color per iteration)



# Knowledge = Neurons' content

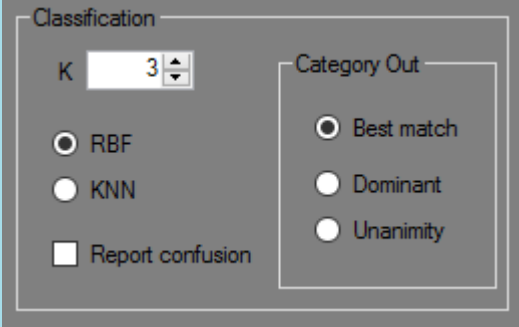
- ▶ Context
- ▶ Category
- ▶ Model (byte array)
- ▶ Influence field
- ▶ Current Min Influence Field
- ▶ Identifier of the pattern which committed the neuron
- ▶ Iteration during which the neuron was committed



# Classification

---

- ▶ Radial Basis Function
- ▶ K\_Nearest Neighbor
- ▶ Both methods uses a K value
- ▶ Consolidation rule in case of uncertainty
- ▶ Report options



Classification

K

RBF  
 KNN  
 Report confusion

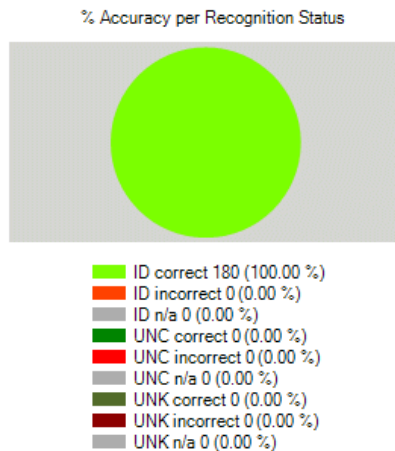
Category Out

Best match  
 Dominant  
 Unanimity

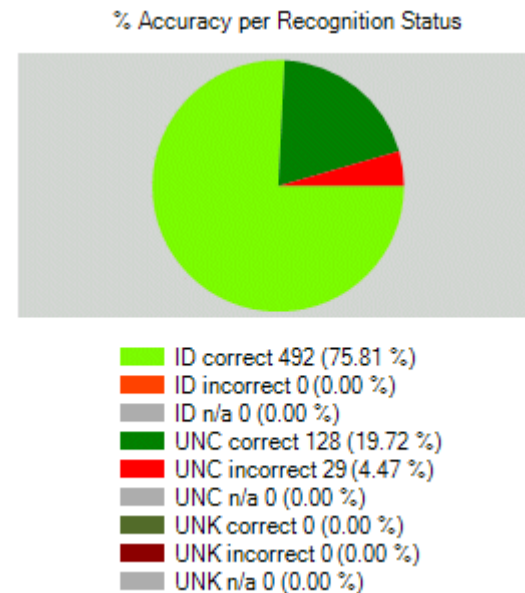
The image shows a software dialog box titled "Classification". It contains a "K" value set to 3, a radio button selection for "RBF" (which is selected), and a checkbox for "Report confusion" which is unchecked. To the right, there is a sub-section titled "Category Out" with three radio button options: "Best match" (selected), "Dominant", and "Unanimity".

# Recognition diagnostics

- ▶ **ID**
  - ▶ Identified, All firing neurons recognize the same category
- ▶ **UNC**
  - ▶ Uncertain, All firing neurons do not recognize the same category
- ▶ **UNK**
  - ▶ Unknown, No firing neurons



- ▶ **Correct**
  - ▶ Best match category = Ground truth category
- ▶ **N/A**
  - ▶ Ground truth category is inexistant; accuracy cannot be evaluated





# Confusion

---

- ▶ Warning that some categories will be difficult to dissociate
  - ▶ Try learning with a smaller Min Influence Field if applicable
  - ▶ Learn more vectors of these categories
  - ▶ Consider a complementary set of feature vectors for these categories
    - ▶ learned under a different new context
    - ▶ Define how to aggregate the responses of the 2 contexts
  - ▶ Consider alternate feature vectors to classify the objects regardless of their categories

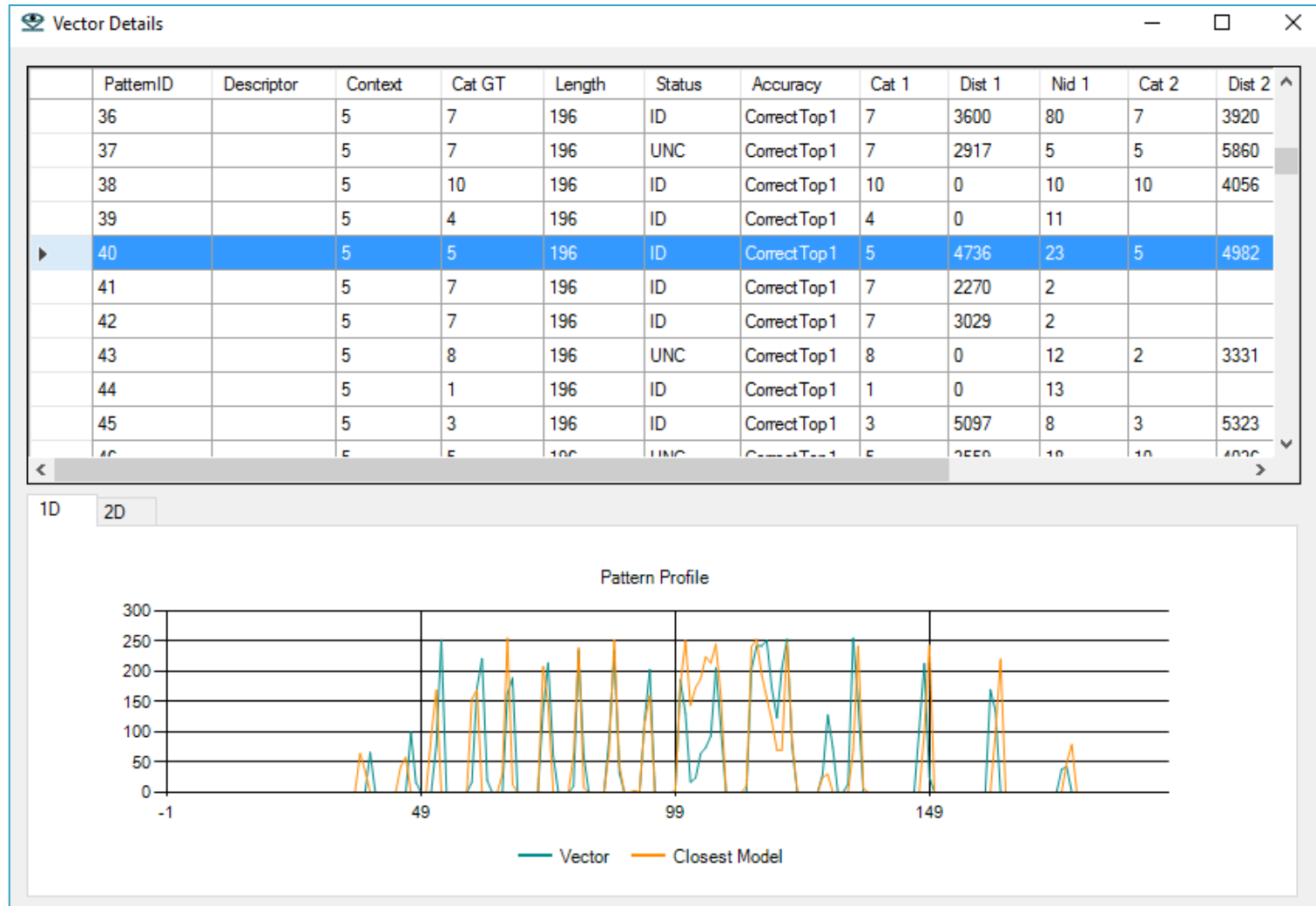
```
Confusion Matrix
0, 0, 0, 0, 0,
0, 54, 0, 0, 0,
0, 0, 34, 0, 0,
0, 0, 0, 79, 0,
0, 0, 0, 0, 13,
```

no confusion

```
Confusion Matrix
0, 0, 0, 0, 0,
0, 1122, 0, 0, 0,
0, 11, 468, 0, 0,
0, 0, 0, 1460, 0,
0, 0, 0, 0, 154,
```

11 cases of confusion:  
GT cat=2 are recognized as cat=1,  
But no GT cat=1 recognized as cat=2

# Recognition Results



# Details

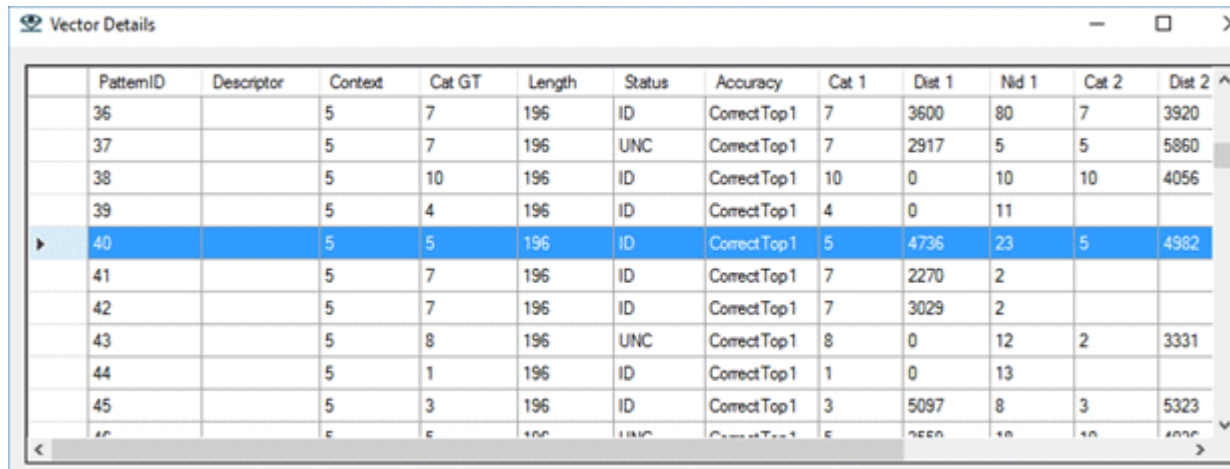
Vector Details



|   | PatternID | Descriptor | Context | Cat GT | Length | Status | Accuracy    | Cat 1 | Dist 1 | Nid 1 | Cat 2 | Dist 2 | ^ |
|---|-----------|------------|---------|--------|--------|--------|-------------|-------|--------|-------|-------|--------|---|
|   | 3207      |            | 3       | 4      | 192    | ID     | CorrectTop1 | 4     | 6642   | 3     |       |        |   |
|   | 3208      |            | 3       | 4      | 192    | ID     | CorrectTop1 | 4     | 7026   | 3     |       |        |   |
|   | 3209      |            | 3       | 3      | 192    | ID     | CorrectTop1 | 3     | 7455   | 8     |       |        |   |
|   | 3210      |            | 3       | 3      | 192    | ID     | CorrectTop1 | 3     | 6768   | 2     | 3     | 7379   |   |
|   | 3211      |            | 3       | 3      | 192    | ID     | CorrectTop1 | 3     | 5738   | 2     | 3     | 6293   |   |
| ▶ | 3212      |            | 3       | 4      | 192    | UNC    | CorrectTop1 | 4     | 6268   | 7     | 3     | 6294   |   |
|   | 3213      |            | 3       | 3      | 192    | UNC    | CorrectTop1 | 3     | 7247   | 2     |       | 8596   |   |
|   | 3214      |            | 3       | 3      | 192    | ID     | CorrectTop1 | 3     | 7741   | 2     |       |        |   |
|   | 3215      |            | 3       | 3      | 192    | ID     | CorrectTop1 | 3     | 6890   | 8     | 3     | 7911   |   |
| * |           |            |         |        |        |        |             |       |        |       |       |        |   |

Recognition of PatternID #3212 is reported as incorrect because the category of the 1<sup>st</sup> firing neuron Cat 1 <> “3”,  
But note that the category of the 2<sup>nd</sup> firing neuron Cat2 = “3”

# RBF versus KNN



| PatternID | Descriptor | Context | Cat GT | Length | Status | Accuracy    | Cat 1 | Dist 1 | Nid 1 | Cat 2 | Dist 2 |
|-----------|------------|---------|--------|--------|--------|-------------|-------|--------|-------|-------|--------|
| 36        |            | 5       | 7      | 196    | ID     | CorrectTop1 | 7     | 3600   | 80    | 7     | 3920   |
| 37        |            | 5       | 7      | 196    | UNC    | CorrectTop1 | 7     | 2917   | 5     | 5     | 5860   |
| 38        |            | 5       | 10     | 196    | ID     | CorrectTop1 | 10    | 0      | 10    | 10    | 4056   |
| 39        |            | 5       | 4      | 196    | ID     | CorrectTop1 | 4     | 0      | 11    |       |        |
| 40        |            | 5       | 5      | 196    | ID     | CorrectTop1 | 5     | 4736   | 23    | 5     | 4982   |
| 41        |            | 5       | 7      | 196    | ID     | CorrectTop1 | 7     | 2270   | 2     |       |        |
| 42        |            | 5       | 7      | 196    | ID     | CorrectTop1 | 7     | 3029   | 2     |       |        |
| 43        |            | 5       | 8      | 196    | UNC    | CorrectTop1 | 8     | 0      | 12    | 2     | 3331   |
| 44        |            | 5       | 1      | 196    | ID     | CorrectTop1 | 1     | 0      | 13    |       |        |
| 45        |            | 5       | 3      | 196    | ID     | CorrectTop1 | 3     | 5097   | 8     | 3     | 5323   |
| 46        |            | 5       | 5      | 196    | UNC    | CorrectTop1 | 5     | 2550   | 10    | 10    | 4000   |

# Methodology

---

# Validation Level 1: on training set

---

- ▶ **Recognize the learning set**
  - ▶ Radial Basis Function
- ▶ **Status of the recognition**
  - ▶ Expect the maximum identified (ID)
  - ▶ Possible uncertainties (UNC)
  - ▶ Expect no unknown (UNK), unless category 0 is a possible Ground Truth category
- ▶ **Accuracy of the recognition**
  - ▶ Expect no Incorrect
  - ▶ Possible n/a if some inputs have no associated Ground truth category

# Validation Level 2: on testing set

---

- ▶ Recognize new vectors with their Ground Truth categories
- ▶ Classification options
  - ▶ Radial Basis Function
  - ▶ K Nearest neighbors
- ▶ Qualify accuracy of the classifier
  - ▶ Correct per single and aggregate categories
  - ▶ If ID\_Incorrect, convert them into UNC by increasing the MINIF. UNC can be trigger the use of a 2<sup>nd</sup> network
- ▶ Qualify the throughput of the classifier
  - ▶ Ratio of positive versus unknown classifications
- ▶ Build a matrix of confusion
  - ▶ Discrepancies between learned and recognized categories