Prevention of occurrence of dead units in self-organizing maps

Student: Bc. Jakub Novák
Supervisor: doc. RNDr. Martin Takáč, PhD.

Contents

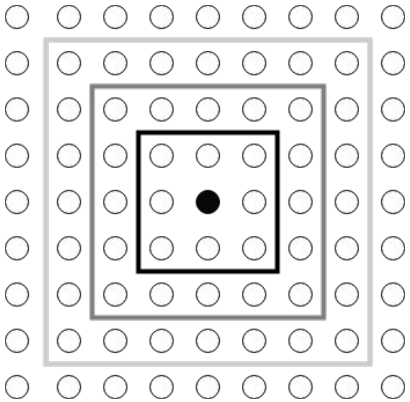
- Goals
- Self-organizing map
- Dead units
- Method
- Results and comparison
- Future work

Goals

- review existing approaches
- suggest SOM modifications
- implement simulations
- analyze to what extent the proposed modifications succeed in eliminating dead units

Self-organizing map

- model of artificial neural network trained using unsupervised learning
- preserves topology
- two-dimensional representation of the input space
- neighbourhood function
  - adapts neighbourhood of winning neuron
  - neighbourhood shrinks with time



(1) Neighbourhood

Self-organizing map

- SOM algorithm has 2 steps
  - competition
  - learning
- competition
  - neurons compete, which one has weights the closest to the input
- learning
  - neurons adapt weights within the neighbourhood
  - SOM parameters update (learning rate, neighbourhood size, ...)

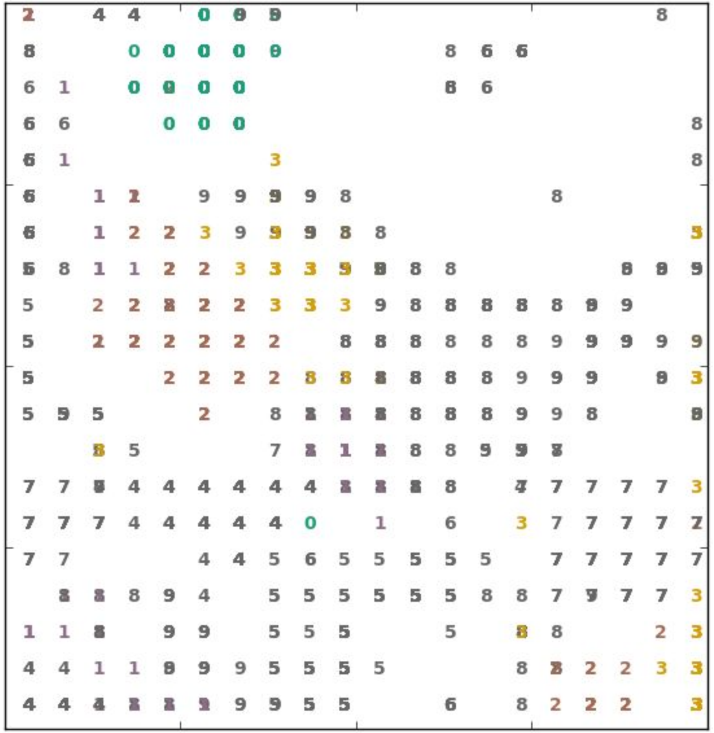
Motivation

- neurons remember information about available data in their weights
- small map forces neurons to represent more different data
- bigger map can represent data more precisely
    - unused neurons, called dead units, might occur

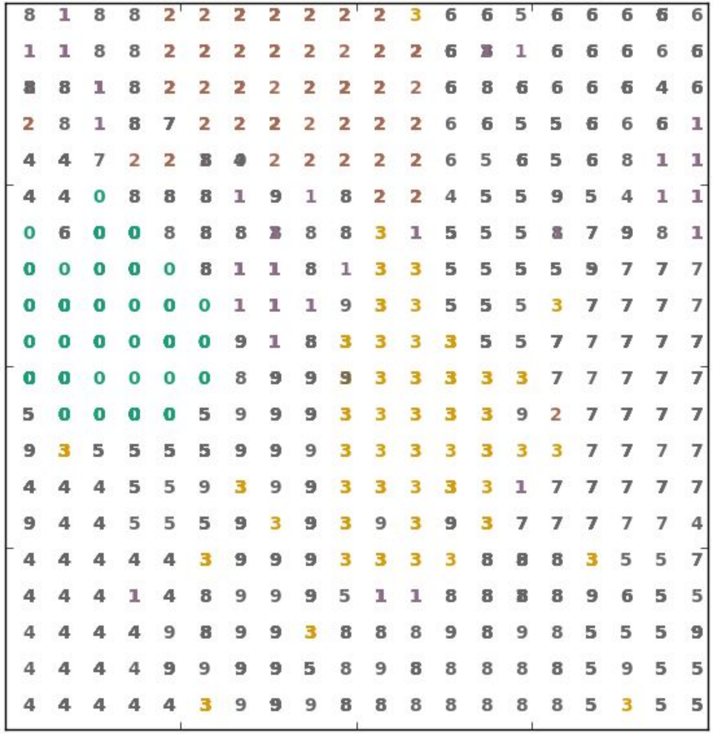
Dead units problem

- common
- as a result, network capacity is not being fully utilized
- usually caused by badly initialized weights in the SOM
    - some neurons have weights far from any data
- rich get richer
- losers adapt too little

Dead units problem



(2) 20x20 SOM with dead units



(3) 20x20 SOM without dead units

General setup

- SOM of size 20x20
- scikit-learn digit dataset
  - 1797 hand written 8x8 digits
- 30 epochs
- Gaussian neighborhood size

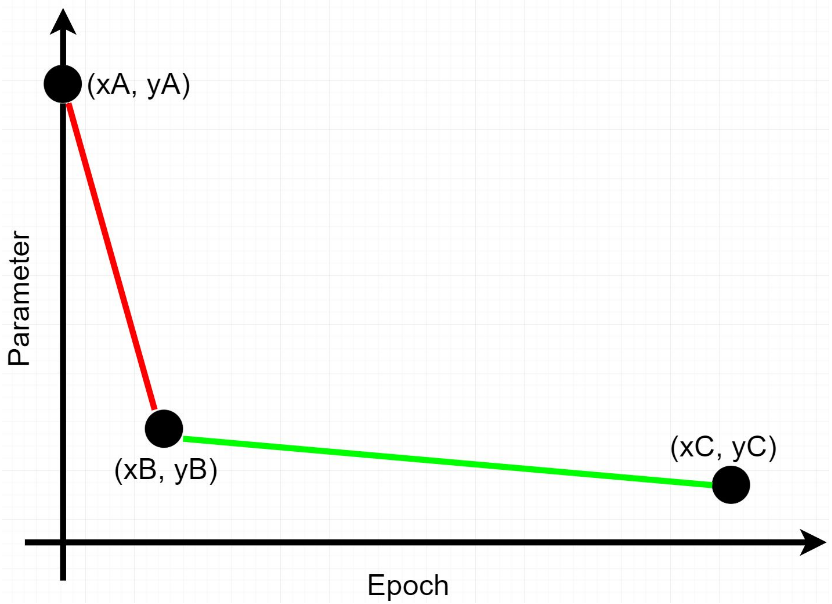
Method

- neighbourhood size and learning rate annealing
- training random dead unit
- training dead units for novel inputs

Neighbourhood size and learning rate annealing

- standard approach
- two phases
  - initial organization phase
    - start off with big neighbourhood size and quickly (within few epochs) get to small neighbourhood size
  - fine-tuning phase
    - fine-tuning neuron weights

  - (xA, yA) - starting point
    (xB, yB) - breaking point
    (xC, yC) - end point



(4) Annealing

Neighbourhood size and learning rate annealing

- explored all combinations for
  - breaking points in 1st, 3rd, 7th and 12th epoch
  - breaking point learning rates: 0.1, 0.2, 0.5
  - initial neighbourhood sizes: 2, 5, 10, 20, 50, 100
- hypothesis
  - starting with larger neighbourhood sizes can help minimize the number of dead units

Neighbourhood size and learning rate annealing

Breaking point alpha: 0.2		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	13.00%	11.25%	9.50%	12.75%	12.00%	11.75%
	3	10.50%	9.00%	13.50%	7.50%	13.25%	11.75%
	7	10.50%	13.00%	7.00%	10.50%	11.30%	10.50%
	12	12.25%	8.75%	9.50%	11.25%	7.75%	9.00%
	20	10.00%	10.50%	9.75%	10.00%	10.75%	9.75%

(5) Percentage of dead units for each parameter combination using first method.

Breaking point alpha: 0.5		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	8.75%	9.50%	9.75%	11.25%	9.75%	9.25%
	3	10.25%	8.50%	11.00%	9.25%	11.00%	11.25%
	7	11.00%	12.00%	11.50%	12.50%	11.50%	10.50%
	12	10.00%	8.50%	9.50%	10.00%	10.75%	10.00%
	20	11.50%	12.00%	11.25%	9.75%	12.25%	12.50%

Training random dead unit

- extra step after each epoch
- randomly choose dead unit
- find closest input datum
- adapt dead units weights and weights of its neighbourhood with this datum

Training random dead unit

- two variants
  - remove dead unit from map of dead units
  - keep dead unit in map of dead units
- hypothesis
  - eliminate dead neurons by forcing them to adapt

Training random dead unit

- keeping the dead unit in map of dead units achieved better results (8.25-13.75% vs 7-13% of dead units)

Breaking point alpha: 0.2		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	11.75%	10.25%	11.50%	10.50%	11.00%	11.50%
	3	10.25%	11.00%	12.25%	9.50%	11.75%	11.25%
	7	11.00%	9.00%	11.50%	10.75%	10.25%	11.50%
	12	9.00%	10.00%	9.00%	10.00%	9.25%	9.25%
	20	11.75%	11.00%	10.25%	9.75%	12.50%	10.25%

(6) Results for variant that removed dead unit from the map of dead units.

Breaking point alpha: 0.2		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	11.75%	9.75%	13.00%	8.00%	11.00%	10.50%
	3	12.75%	11.50%	11.50%	11.25%	12.25%	10.00%
	7	13.00%	10.25%	10.00%	13.00%	13.25%	12.50%
	12	7.00%	11.25%	11.00%	10.00%	12.00%	9.25%
	20	11.00%	12.50%	11.25%	11.00%	9.50%	12.50%

(7) Results for variant that kept dead unit in the map of dead units.

Training dead units for novel inputs

- are neurons weights similar enough to the input datum?
- threshold variable
  - some digits are not clearly separable

- choosing threshold
  - calculate mean Euclidean distance for inputs of the same category
  - calculate cross-category Euclidean distances

Training dead units for novel inputs

- two variants
  - dead units are adapted with learning rate according to annealing scheme
  - dead units are adapted with constant learning rate (1.0)
- hypothesis
  - potential better clusterization
  - novelty inputs mapped to unused neurons

Training random dead unit

- using decayed learning rate according to annealing scheme achieved slightly better results

Breaking point alpha: 0.1		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	11.00%	11.50%	8.50%	10.00%	10.50%	11.50%
	3	13.25%	13.00%	10.50%	10.50%	10.75%	9.25%
	7	10.00%	7.50%	10.50%	10.50%	9.75%	11.00%
	12	13.25%	9.50%	12.25%	8.00%	11.50%	10.75%
	20	10.50%	12.75%	9.00%	12.25%	12.25%	12.75%

(8) Results for variant that used decayed learning rate.

Breaking point alpha: 0.1		Initial sigma					
		2.0	5.0	10.0	20.0	50.0	100.0
Breaking point	1	13.80%	13.30%	13.50%	14.30%	15.00%	9.80%
	3	11.50%	12.00%	13.30%	15.30%	12.30%	12.50%
	7	10.80%	12.30%	12.00%	14.00%	10.00%	13.80%
	12	10.30%	11.30%	10.00%	10.80%	13.50%	15.50%
	20	11.30%	9.00%	9.80%	14.00%	13.30%	15.00%

(9) Results for variant that used constant 1.0 learning rate.

Across method comparison

1. Big neighborhood size (baseline)

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	13.50%	11.50%	10.50%	11.75%	12.25%	10.50%
3	12.25%	10.75%	12.25%	12.75%	11.50%	11.75%
7	10.30%	11.30%	9.50%	10.00%	11.50%	12.80%
12	8.75%	8.25%	11.75%	9.75%	10.00%	10.25%
20	11.25%	10.50%	11.00%	10.50%	11.25%	11.25%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	13.00%	11.25%	9.50%	12.75%	12.00%	11.75%
3	10.50%	9.00%	13.50%	7.50%	13.25%	11.75%
7	10.50%	13.00%	7.00%	10.50%	11.30%	10.50%
12	12.25%	8.75%	9.50%	11.25%	7.75%	9.00%
20	10.00%	10.50%	9.75%	10.00%	10.75%	9.75%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	8.75%	9.50%	9.75%	11.25%	9.75%	9.25%
3	10.25%	8.50%	11.00%	9.25%	11.00%	11.25%
7	11.00%	12.00%	11.50%	12.50%	11.50%	10.50%
12	10.00%	8.50%	9.50%	10.00%	10.75%	10.00%
20	11.50%	12.00%	11.25%	9.75%	12.25%	12.50%

2. Remove DU from dead_units list after adapting it

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	13.75%	12.00%	12.75%	12.50%	12.50%	10.00%
3	10.00%	10.75%	12.25%	9.25%	12.50%	10.75%
7	11.25%	10.25%	9.25%	10.50%	12.75%	11.50%
12	10.75%	11.25%	8.25%	9.25%	12.25%	11.50%
20	8.25%	10.50%	12.25%	10.50%	10.00%	11.50%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.75%	10.25%	11.50%	10.50%	11.00%	11.50%
3	10.25%	11.00%	12.25%	9.50%	11.75%	11.25%
7	11.00%	9.00%	11.50%	10.75%	10.25%	11.50%
12	9.00%	10.00%	9.00%	10.00%	9.25%	9.25%
20	11.75%	11.00%	10.25%	9.75%	12.50%	10.25%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	9.50%	10.25%	0.115	11.25%	10.75%	10.25%
3	10.25%	11.00%	10.50%	11.75%	12.50%	9.50%
7	10.25%	11.75%	12.00%	11.00%	10.00%	11.50%
12	11.75%	10.75%	9.50%	10.75%	10.25%	10.75%
20	11.00%	10.75%	9.00%	13.75%	11.00%	11.50%

2. Keep DU in dead_units list after adapting it

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.75%	11.50%	12.25%	10.75%	11.00%	10.75%
3	10.50%	7.50%	9.00%	10.50%	10.50%	12.50%
7	10.00%	10.75%	9.75%	10.50%	12.75%	11.75%
12	9.75%	10.50%	10.75%	9.75%	10.50%	12.50%
20	10.00%	9.25%	10.50%	12.50%	13.50%	11.50%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.75%	9.75%	13.00%	8.00%	11.00%	10.50%
3	12.75%	11.50%	11.50%	11.25%	12.25%	10.00%
7	13.00%	10.25%	10.00%	13.00%	13.25%	12.50%
12	7.00%	11.25%	11.00%	10.00%	12.00%	9.25%
20	11.00%	12.50%	11.25%	11.00%	9.50%	12.50%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	10.50%	12.00%	9.00%	11.75%	8.75%	10.25%
3	12.25%	11.25%	10.25%	12.25%	13.00%	9.50%
7	10.75%	14.00%	10.75%	9.50%	11.00%	10.75%
12	8.25%	10.50%	11.75%	10.25%	12.25%	10.75%
20	10.75%	10.00%	11.00%	11.00%	13.50%	11.00%

3. Threshold = 0.4, Learning rate depends on decay

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.00%	11.50%	8.50%	10.00%	10.50%	11.50%
3	13.25%	13.00%	10.50%	10.50%	10.75%	9.25%
7	10.00%	7.50%	10.50%	10.50%	9.75%	11.00%
12	13.25%	9.50%	12.25%	8.00%	11.50%	10.75%
20	10.50%	12.75%	9.00%	12.25%	12.25%	12.75%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	12.00%	10.75%	11.50%	11.00%	9.25%	13.00%
3	8.75%	11.50%	11.50%	10.00%	12.00%	10.25%
7	11.75%	12.25%	12.50%	8.75%	11.50%	11.75%
12	10.25%	10.25%	11.75%	12.00%	9.00%	11.00%
20	10.50%	12.25%	12.00%	12.25%	12.00%	9.00%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.50%	9.50%	8.75%	9.00%	13.25%	11.25%
3	12.00%	9.00%	10.75%	12.75%	13.50%	10.75%
7	11.50%	9.00%	13.50%	10.75%	12.75%	10.00%
12	8.25%	11.50%	10.00%	11.00%	10.25%	10.00%
20	11.75%	9.25%	12.00%	10.50%	10.50%	12.00%

3. Threshold = 0.5, Learning rate depends on decay

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	12.00%	11.75%	11.25%	11.00%	12.25%	11.25%
3	11.75%	12.00%	9.25%	9.75%	10.50%	8.25%
7	12.00%	9.25%	9.75%	11.25%	12.25%	11.75%
12	9.50%	8.75%	13.50%	10.00%	11.50%	12.25%
20	12.00%	10.50%	9.50%	11.25%	11.25%	12.75%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	10.25%	12.75%	12.50%	12.50%	13.25%	11.25%
3	8.50%	10.75%	10.50%	10.00%	9.50%	10.00%
7	11.50%	10.25%	10.50%	11.25%	9.25%	10.75%
12	10.25%	11.50%	9.50%	11.25%	11.50%	11.25%
20	11.00%	10.25%	12.25%	11.00%	12.00%	12.75%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	11.50%	11.25%	11.75%	13.25%	9.50%	12.00%
3	10.75%	12.25%	11.75%	10.00%	12.50%	8.50%
7	9.25%	12.50%	9.50%	10.75%	11.50%	11.00%
12	10.50%	10.75%	10.75%	10.75%	7.50%	10.25%
20	12.50%	9.00%	13.00%	9.75%	11.00%	11.25%

3. Threshold = 0.6, Learning rate depends on decay

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	12.75%	11.50%	9.75%	11.75%	13.50%	11.50%
3	13.25%	13.50%	10.75%	12.00%	10.50%	10.00%
7	9.00%	10.25%	9.50%	11.00%	11.75%	12.50%
12	9.25%	10.25%	11.25%	10.00%	10.00%	11.50%
20	9.25%	12.25%	10.50%	12.25%	12.25%	12.25%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	13.50%	11.50%	11.25%	10.50%	13.00%	11.00%
3	11.00%	12.25%	11.00%	10.75%	12.50%	10.00%
7	8.75%	12.00%	9.25%	11.25%	12.00%	11.75%
12	12.25%	11.25%	10.75%	12.25%	13.00%	11.50%
20	10.00%	10.50%	9.75%	8.50%	12.25%	8.75%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	12.00%	9.50%	11.00%	10.00%	11.00%	10.25%
3	9.00%	11.50%	10.25%	12.00%	11.00%	12.25%
7	11.50%	12.25%	11.50%	9.75%	12.50%	11.00%
12	9.00%	11.00%	11.50%	12.25%	9.75%	10.25%
20	8.75%	11.00%	11.00%	12.25%	10.75%	11.00%

3. Threshold = 0.4, Learning rate = 1.0

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	13.80%	13.30%	13.50%	14.30%	15.00%	9.80%
3	11.50%	12.00%	13.30%	15.30%	12.30%	12.50%
7	10.80%	12.30%	12.00%	14.00%	10.00%	13.80%
12	10.30%	11.30%	10.00%	10.80%	13.50%	15.50%
20	11.30%	9.00%	9.80%	14.00%	13.30%	15.00%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	10.80%	16.50%	14.00%	13.50%	13.30%	13.00%
3	10.30%	11.00%	11.50%	13.50%	13.50%	11.50%
7	10.00%	12.00%	11.50%	12.30%	10.80%	14.30%
12	9.30%	10.80%	9.80%	11.80%	11.00%	14.00%
20	9.50%	10.00%	11.00%	10.80%	13.00%	14.00%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	10.80%	11.30%	11.80%	9.30%	11.00%	9.80%
3	10.00%	10.50%	12.00%	10.30%	11.50%	11.80%
7	12.30%	9.00%	10.30%	13.80%	10.50%	11.00%
12	12.30%	9.80%	11.50%	10.30%	9.30%	13.30%
20	12.50%	10.00%	9.80%	9.30%	12.30%	12.50%

3. Threshold = 0.5, Learning rate = 1.0

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	45.00%	50.80%	15.50%	43.80%	41.50%	46.50%
3	13.30%	42.00%	45.30%	46.30%	45.00%	39.30%
7	11.50%	10.30%	12.30%	43.50%	47.30%	18.00%
12	12.30%	14.00%	11.30%	15.50%	50.00%	43.50%
20	13.50%	12.80%	14.80%	12.30%	15.00%	48.50%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	43.50%	44.50%	42.80%	49.30%	45.80%	42.30%
3	13.30%	40.00%	45.80%	37.30%	46.00%	42.80%
7	10.00%	9.30%	8.80%	37.50%	43.00%	44.80%
12	11.80%	13.00%	9.50%	12.30%	39.80%	42.30%
20	13.30%	8.80%	13.30%	12.50%	33.80%	51.30%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	33.80%	41.00%	36.50%	16.00%	16.80%	34.30%
3	11.50%	38.00%	16.30%	33.00%	34.50%	33.00%
7	10.80%	12.50%	11.80%	18.50%	17.80%	27.30%
12	11.30%	10.30%	11.80%	12.00%	40.30%	38.00%
20	13.00%	10.80%	11.80%	12.80%	39.50%	35.80%

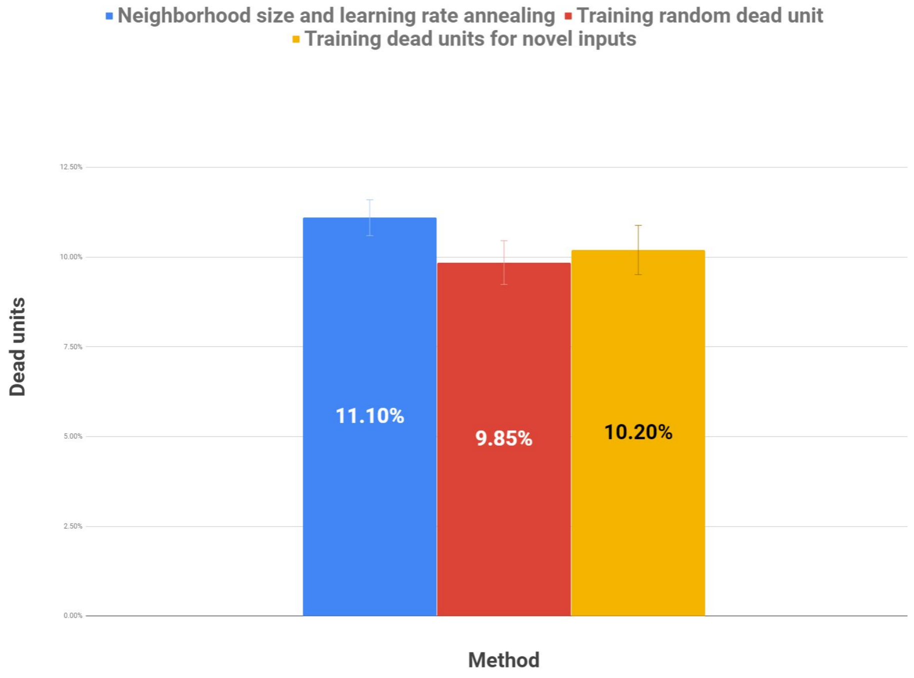
3. Threshold = 0.6, Learning rate = 1.0

Breaking point alpha: 0.1	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	23.80%	25.00%	27.00%	22.80%	24.80%	26.80%
3	32.80%	32.30%	34.30%	26.50%	25.30%	26.00%
7	41.80%	35.50%	58.00%	25.80%	25.00%	26.80%
12	46.00%	12.30%	15.00%	48.80%	31.50%	31.30%
20	37.00%	23.50%	29.00%	65.00%	42.00%	41.50%

Breaking point alpha: 0.2	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	31.30%	26.50%	26.50%	26.30%	25.00%	25.00%
3	36.00%	35.50%	29.80%	25.80%	25.80%	20.30%
7	39.30%	11.80%	14.00%	34.80%	26.30%	26.30%
12	36.00%	15.00%	12.80%	47.50%	30.50%	26.50%
20	27.30%	16.00%	35.00%	63.30%	54.30%	43.50%

Breaking point alpha: 0.5	Initial sigma 2.0	5.0	10.0	20.0	50.0	100.0
1	25.80%	26.30%	22.80%	22.00%	26.00%	22.50%
3	31.80%	33.80%	40.00%	25.30%	28.30%	21.50%
7	34.00%	11.30%	46.50%	32.80%	25.30%	31.30%
12	35.00%	15.30%	13.30%	37.80%	29.30%	18.80%
20	34.00%	23.30%	18.80%	61.50%	50.30%	53.30%

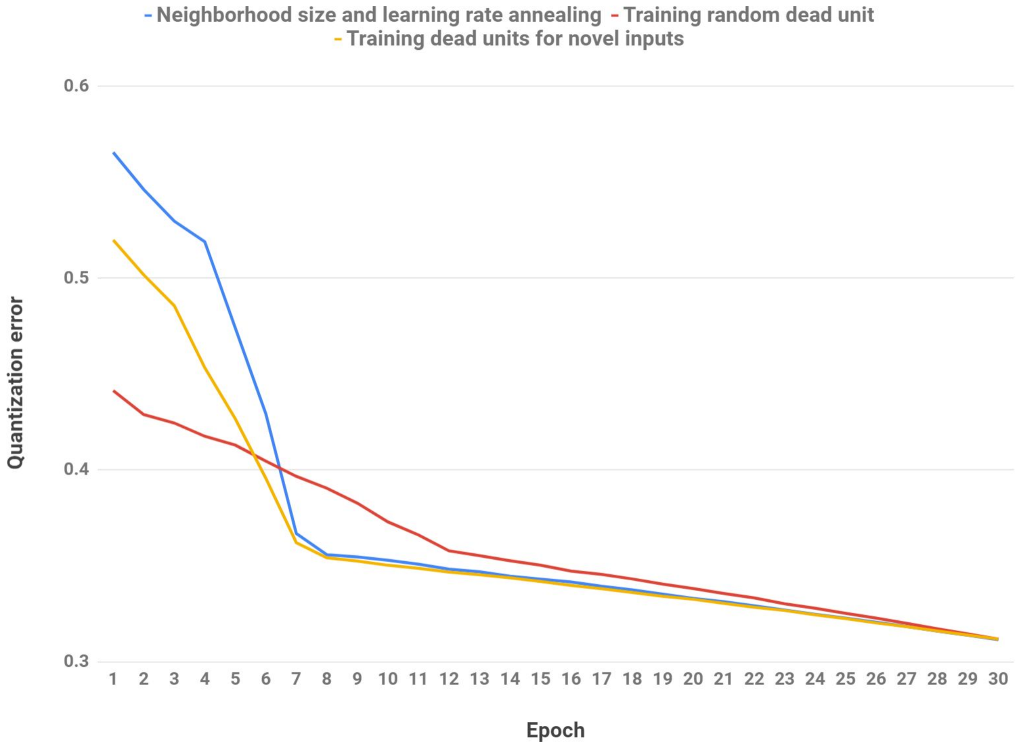
Across method comparison



(10) Average percentage of dead units across methods.

(11) Average quantization error across methods.

Across method comparison



(12) Quantization error convergence across methods.

Across method comparison

- what could be improved based on results
- first method
  - values of neighbourhood size seems to depend on the size of SOM
- second method
  - no significant dependencies
- third method
  - not reliable in current state
  - correct threshold is extremely data dependent

Across method comparison

- third method - bonus experiment
  - for each neuron, remember how much training it got
  - if threshold not satisfied, choose neuron with least training received
  - early results - not good for SOM
    - under-trained neurons disrupt clustering property of SOM

Future work

- different SOM sizes
  - with more neurons than training examples
- train on more datasets
- finer exploration of threshold variable
- criterion function

Thank you for your attention

Questions

- Akým spôsobom získané výsledky závisia od konkrétneho datasetu a konkrétnej veľkosti mapy?
    - priamo
    - problém separability
    - neuróny nútené reprezentovať viac kategórií dát
    - väčšia mapa by nám mohla pomôcť vidieť väčšie rozdiely medzi navrhnutými metódami
- Ktoré parametre by bolo treba upraviť, keby bola SOM oveľa väčšia?
    - veľkosť okolia v štartovacom bode
    - veľkosť okolia v “breaking” bode
    - ostatné parametre nie sú závislé od veľkosti SOM

Questions

- Aký všeobecný záver resp. odporúčania by ste vyvodili?
  - vyskúšať viac veľkostí SOM
  - vyskúšať viac rôznych datasetov