



METODY INŻYNIERII WIEDZY

METHODS OF KNOWLEDGE ENGINEERING

Samoorganizujące się mapy Kohonena

Self-Organizing Maps - SOM



Adrian Horzyk
horzyk@agh.edu.pl



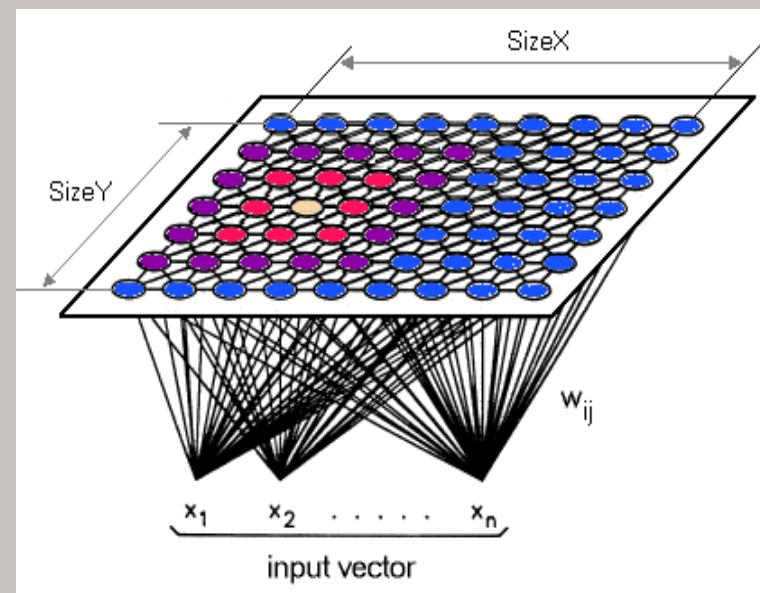
Akademia Górniczo-Hutnicza
w Krakowie

SAMO-ORGANIZUJĄCE SIĘ MAPY



Sieć Kohonena SOM (wymyślone w 1982 r. przez Teuvo Kohonena):

- to sieć o regularnej strukturze składającej się z węzłów, które reprezentują podobne obiekty z przestrzeni wejściowej;
- reprezentująca dane wielowymiarowe w przestrzeni o mniejszym wymiarze (zwykle 2D lub 3D), co nie zawsze umożliwia reprezentację grup podobnych obiektów przez sąsiednie węzły;
- adaptowana z wykorzystaniem nienadzorowanego uczenia konkurencyjnego (*unsupervised competitive learning*), czyli uczenie bez nauczyciela (wskazania celu nauki, tj. klasy, wartości wyjściowych) w taki sposób, iż węzły konkurują pomiędzy sobą o reprezentację danych wejściowych;
- sąsiedztwo węzłów odgrywa istotne znaczenie w trakcie adaptacji sieci, gdyż umożliwia również adaptację węzłów bliskich zwycięzcy prowadząc do automatycznej reorganizacji reprezentacji obiektów przez sieć Kohonena w trakcie procesu uczenia;
- w niewielkim stopniu przypominają sieci neuronowe, gdyż wejścia do węzłów tej sieci nie są ważone, lecz porównywane do wektorów wagowych na podstawie odległości (zwykle Euklidesa);
- wyznaczają średnią z reprezentowanych wzorców w każdym węźle sieci zapisaną w wektorze wag, gdzie reprezentowane są w węźle te wzorce, które powodują iż dany węzeł staje się zwycięzcą (jest najbliższej).



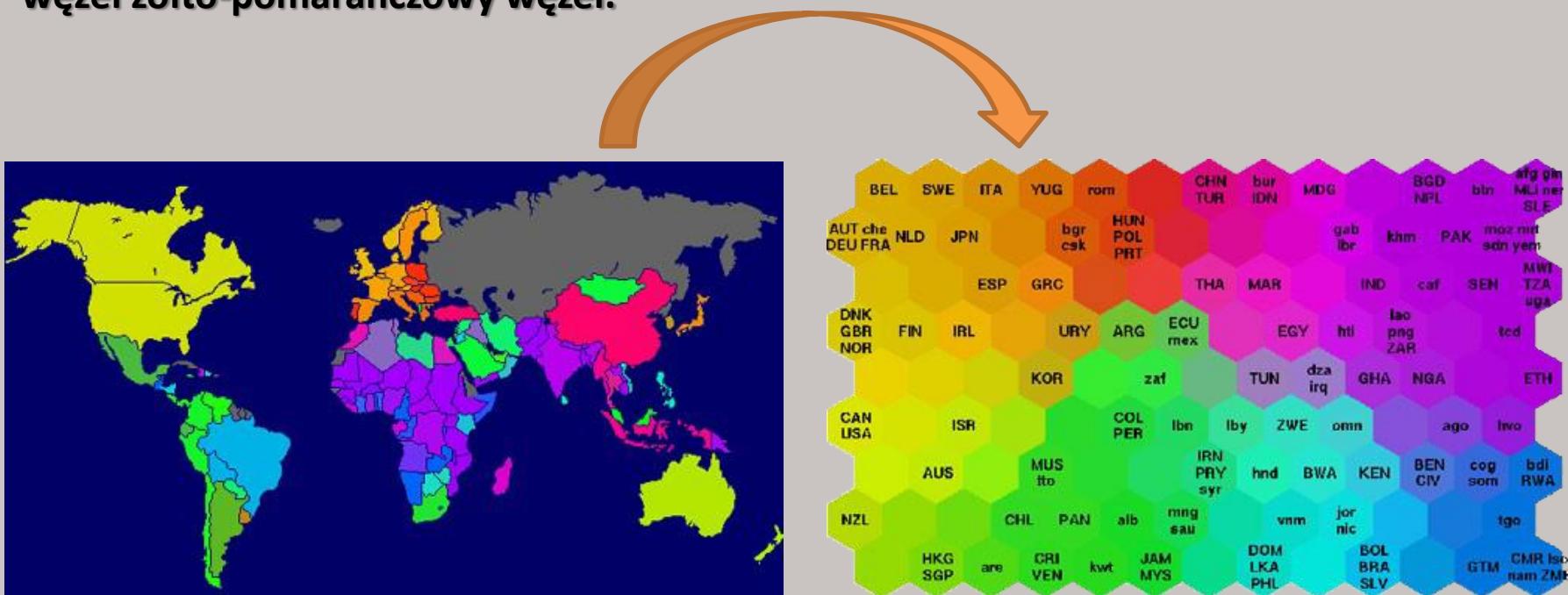


SOM PRZYKŁAD REPREZENTACJI MAPY

Mapa przedstawiająca jakość życia w różnych krajach świata wykorzystana do adaptacji sieci SOM dała następujący rezultat agregując reprezentację różnych krajów oraz w ich sąsiedztwie umieszczając inne kraje o podobnej jakości życia, np.:

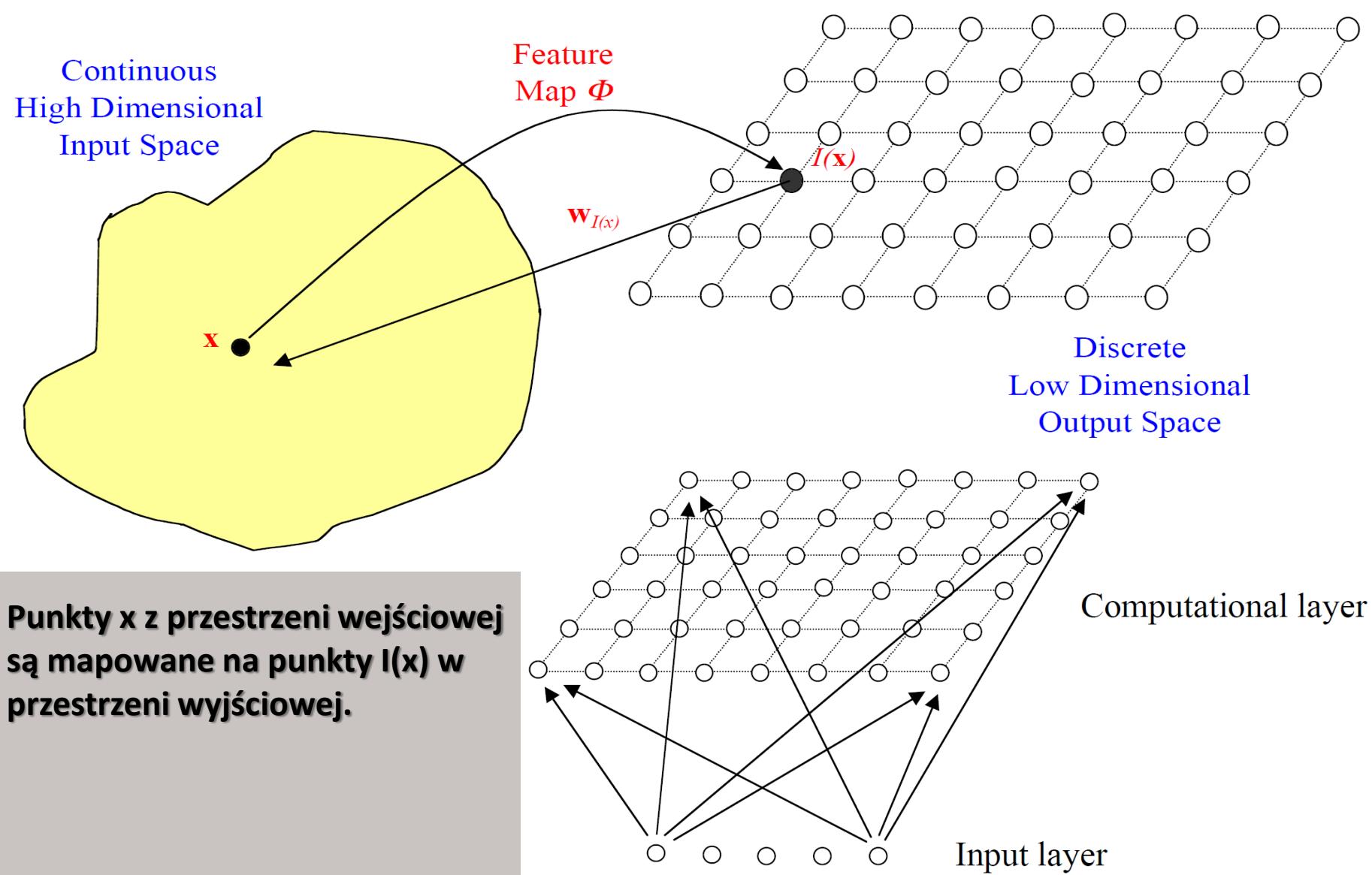
Polska (POL) jest reprezentowana przez ten sam węzeł, co Węgry (HUN) oraz Portugalia (PRT), z lewej strony widzimy bliski węzeł reprezentujący Czechy (CSK).

Kraje wysoko rozwinięte o wyższej stopie życia, tj. USA i Kanada widzimy na żółto, a kraje Europy zachodniej, tj. Niemcy (DEU) czy Francja (FRA) są reprezentowane przez jeden węzeł żółto-pomarańczowy węzeł.





TWORZENIE MAPY DLA DANYCH



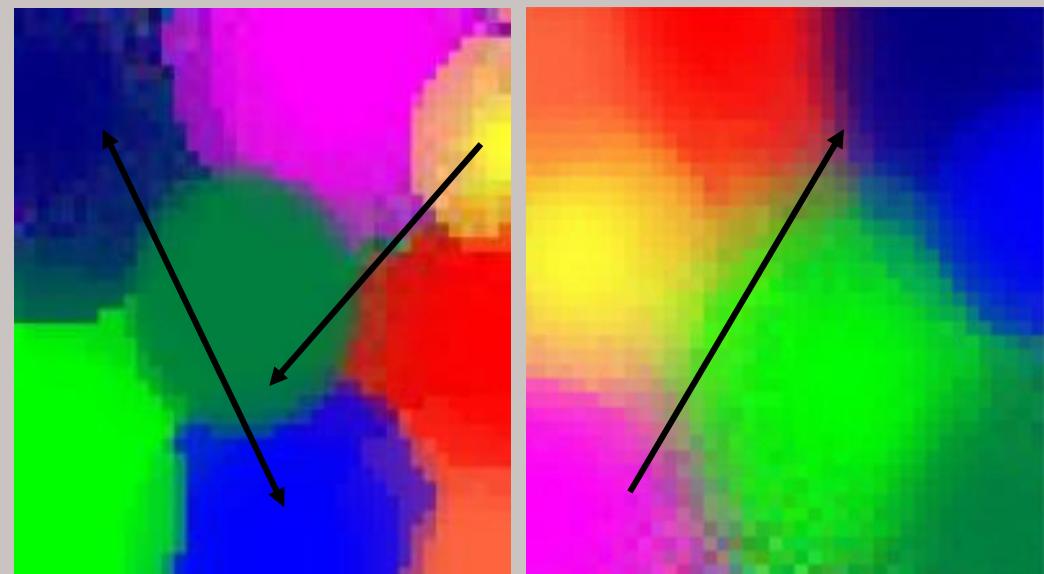
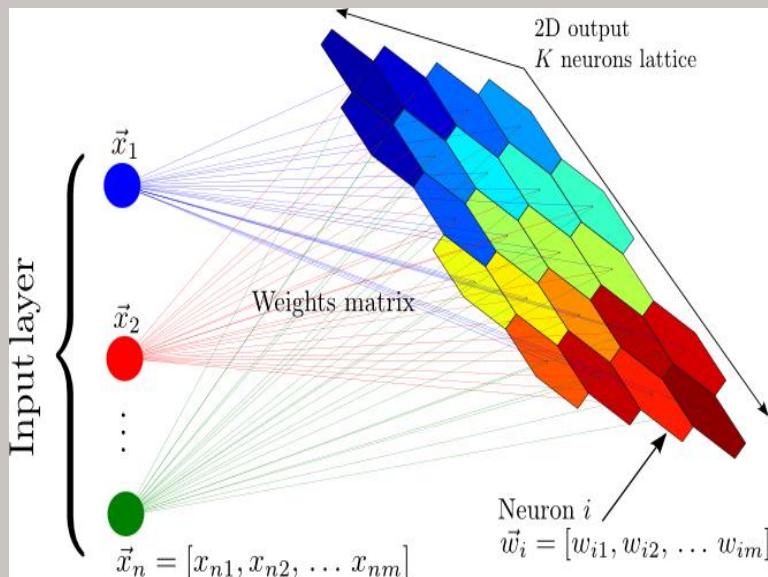


SAMO-ORGANIZUJĄCE SIE MAPY

Próbując odwzorować przestrzeń 3D kolorów RGB w przestrzeni 2D nie jesteśmy w stanie wszystkich podobnych kolorów reprezentować obok siebie:

- **Żółty powinien być pomiędzy czerwonym i zielonym.**
- **Różowy powinien być pomiędzy czerwonym a niebieskim.**
- **Błękitny powinien być pomiędzy zielonym i niebieskim.**

Jak widać na rysunkach poniżej, nie sposób idealnie rzutować przestrzeni 3D na 2D!

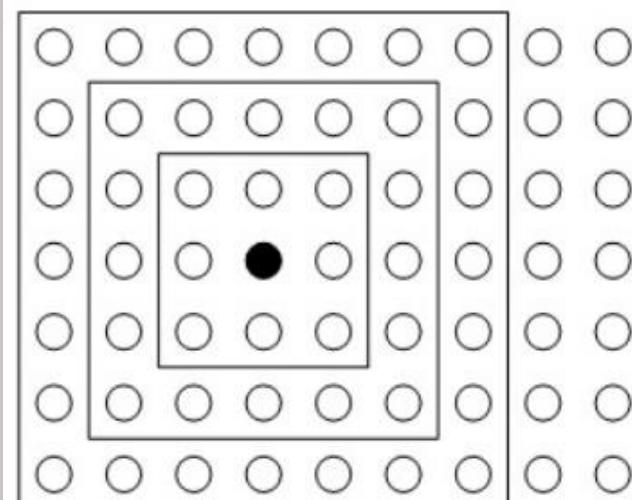




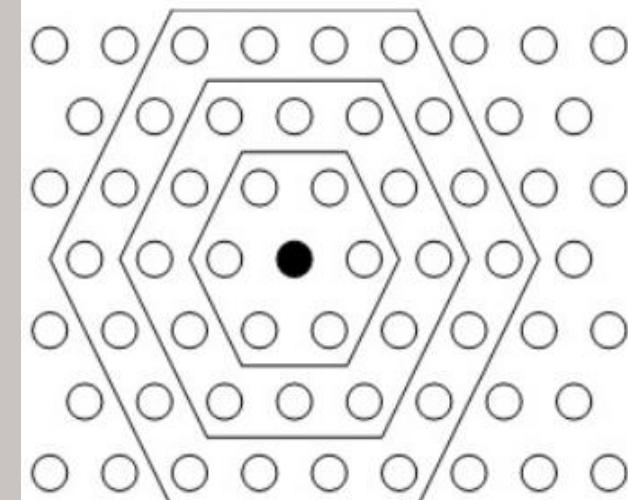
RYWALIZACJA POMIĘDZY WEZŁAMI

Sieć SOM na podstawie podobieństwa wzorców wejściowych aktualizuje wagi węzłów sieci na zasadach rywalizacji (competitive learning) w taki sposób, iż węzeł reprezentujący wartość najbliższą danym wejściowym (najbardziej podobny) w sensie przyjętej metryki (np. Euklidesa) staje się zwycięzcą (winner). Wagi węzła zwycięzcy zostają najmocniej zaktualizowane w kierunku dawnych wejściowych (zmniejszenie odległości) w porównaniu z sąsiednimi wzorcami, których wagi aktualizujemy w mniejszym stopniu, tym mniej im dalej znajdują się od zwycięzcy.

Dzięki modyfikacjom wag sąsiednich węzłów uzyskujemy możliwość reprezentacji podobnych grup wzorców przez sąsiednie węzły sieci SOM:



Siatka prostokątna



Siatka heksagonalna

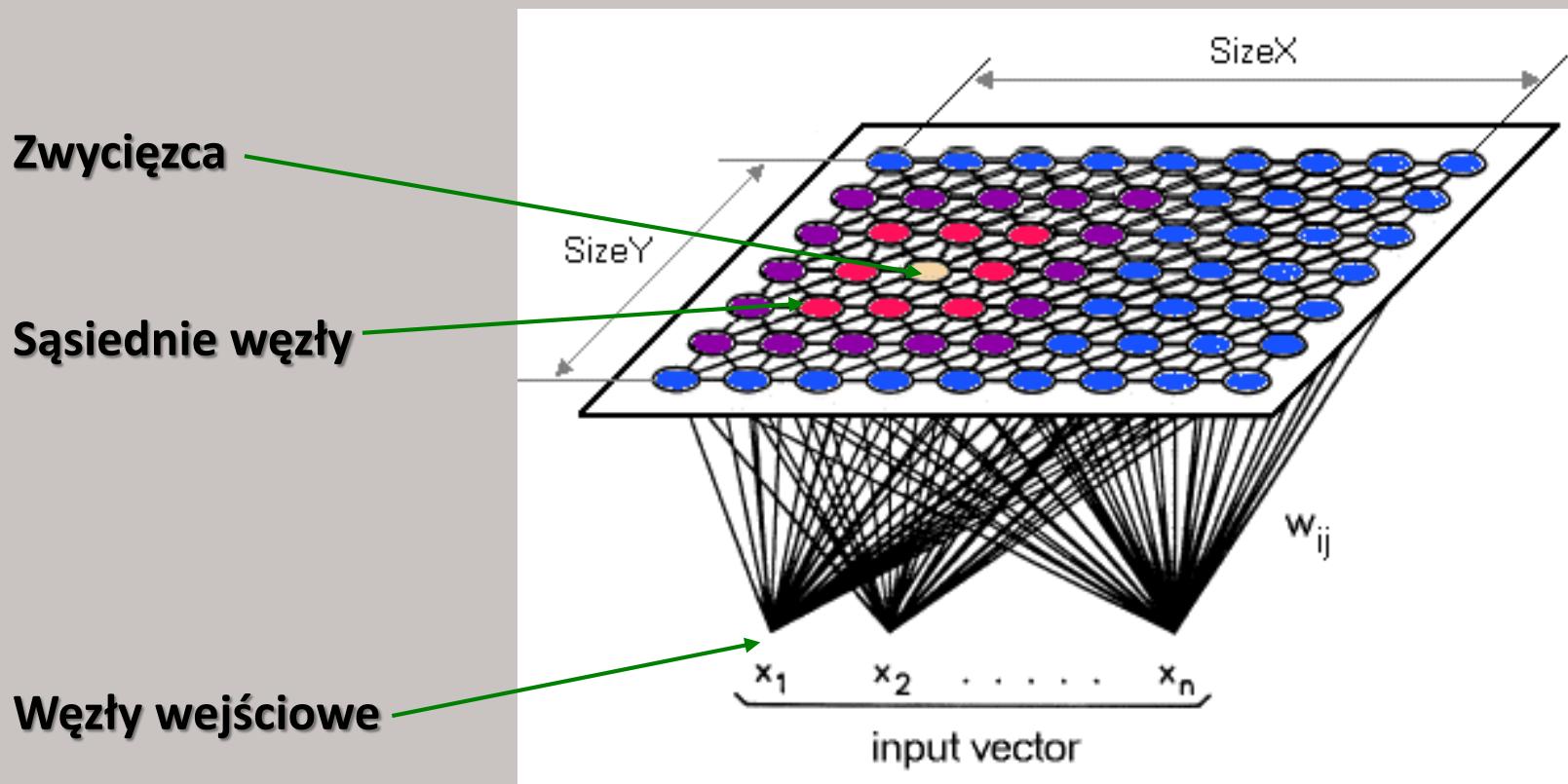


RYWALIZACJA POMIĘDZY WĘZŁAMI

Rozważamy wzorce uczące w postaci wektorów $X_k = \{x_1, x_2, \dots, x_n\}$, których elementy są reprezentowane przez n węzłów wejściowych sieci SOM.

Każdy węzeł wejściowy powiązany jest z każdym węzłem na mapie wyjściowej.

Wagi inicjalizujemy wartościami losowymi, tak żeby węzły reagowały najmocniej na różne kombinacje danych wejściowych $X_k = \{x_1, x_2, \dots, x_n\}$.





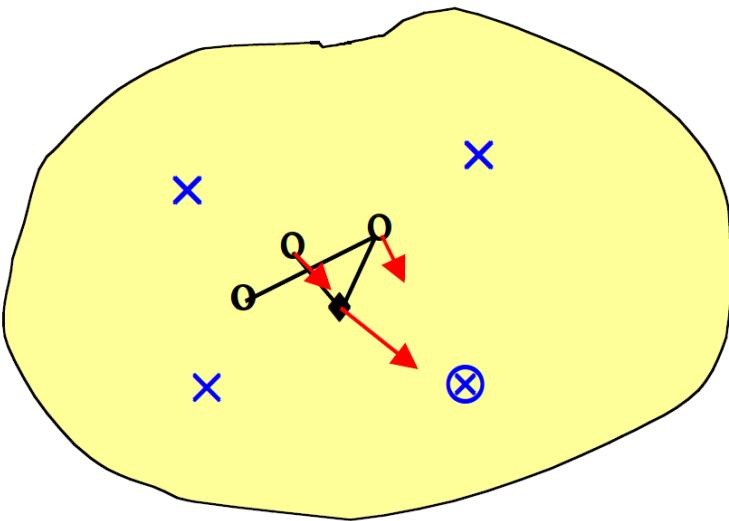
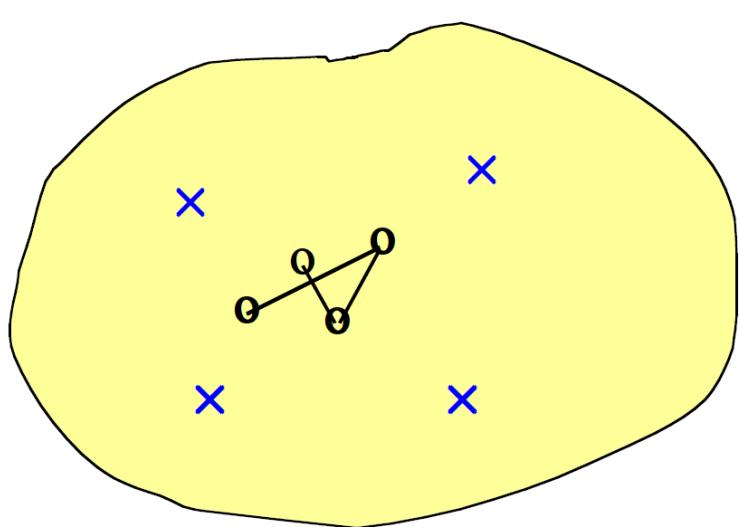
ALGORYTM UCZENIA SIECI SOM

1. Zbuduj wyjściową mapę węzłów (zwykle 2D lub 3D) i zdefiniuj relację sąsiedztwa (kwadratową, heksagonalną, gwiaździstą, promienistą, romboidalną).
2. Inicjalizuj wagi każdego węzła niewielkimi liczbami losowymi różnymi od 0.
3. Bierz kolejne lub losowe wektory $X_k = \{x_1, x_2, \dots, x_n\}$ ze zbioru uczącego X_1, \dots, X_K .
4. Oblicz wartość wyjściową wszystkich węzłów mapy sieci SOM na podstawie wybranej odległości (zwykle Euklidesa, ale można wykorzystać też odległość Manhattan, Mahalanobis itd.) wg poniższego wzoru, wyznaczającego odległość wzorca wejściowego do wektora wag:

$$d(X_k, W_{i,j}(t)) = \sqrt{\sum_{i=1}^I \sum_{j=1}^J (X_k - W_{i,j}(t))^2}$$

5. Określ, który węzeł jest najbliższym wzorcowi wejściowemu $X_k = \{x_1, x_2, \dots, x_n\}$?
6. Aktualizuj wagi zwycięskiego węzła w taki sposób, żeby lepiej odwzorowywały wektor wejściowy (zbliżyły się do niego).
7. Aktualizuj wagi sąsiednich węzłów zwycięzcy w podobny sposób, lecz z malejącym współczynnikiem, tj. im dalej znajdują się od zwycięskiego węzła tym mniej aktualizują swoje wagi.
8. Wróć do kroku 3 dopóki wszystkie wzorce uczące nie zostaną reprezentowane w sieci z wystarczającą dokładnością, co oznacza konieczność wielokrotnego prezentowania wszystkich wzorców uczących sieci.

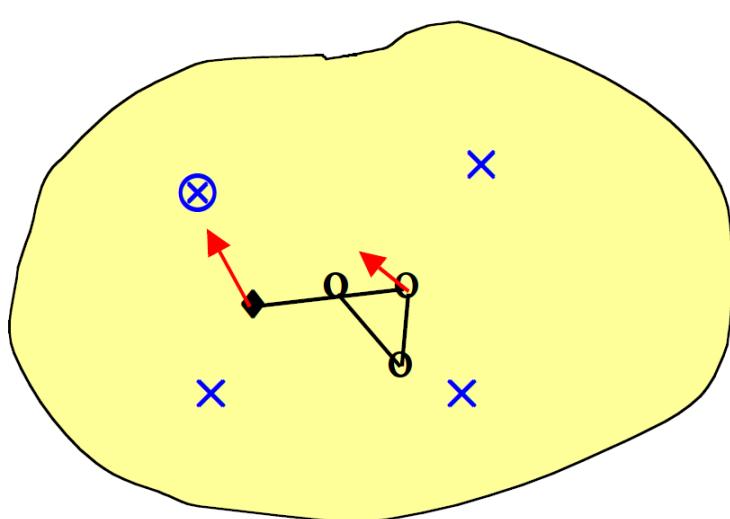
PROCES SAMOORGANIZACJI W SOM



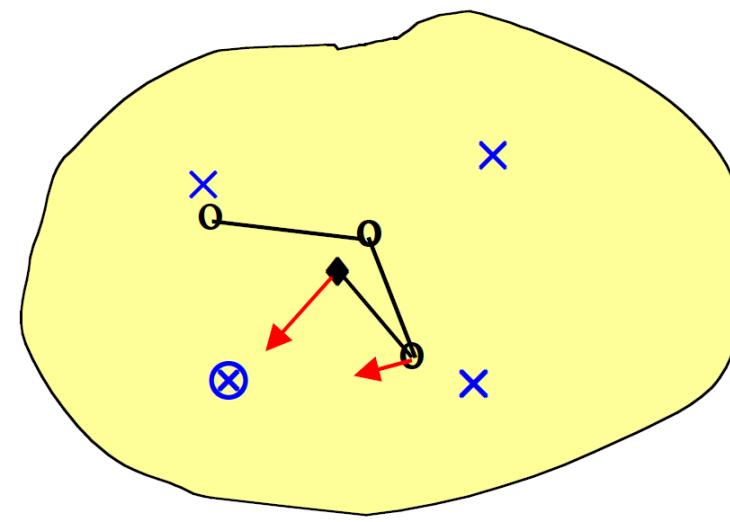
Założymy, że mamy cztery punkty danych (krzyżyki) w naszej ciągłej przestrzeni wejściowej 2D i chcemy je mapować na cztery punkty w dyskretny sposób do 1D przestrzeni wyjściowej. Mapujemy węzły wyjściowe w punkty przestrzeni wejściowej (kółeczka). Losowy początkowe wagi rozpoczynają losowo koła pozycje w środku obszaru wprowadzania.

Losowo wybieramy jeden z punktów danych na trening (krzyżyk w kółku). Najbliższy punkt wyjściowy reprezentuje zwycięski węzeł (pełny diament/rąb). Zwycięski węzeł przesuwamy w kierunku punktu danych mocniej (dłuższa strzałka), a dwa sąsiednie węzły przesuwane są o mniejszą odległość (krótsze strzałki).

PROCES SAMOORGANIZACJI W SOM



Następnie losowo wybieramy inny punkt danych na trening (krzyżyk w kółku). Najbliższy punkt wyjściowy daje nam nowy zwycięski węzeł (pełny diament/rąb). Zwycięski węzeł znowu przesuwa się w kierunku punktu danych z większym krokiem, a jeden sąsiadujący węzeł przesuwa się z mniejszym krokiem.



Kontynuujemy losowe wybieranie punktów danych na trening (krzyżyk w kółku). Każdy wygrywający węzeł przesuwa się w kierunku punktu danych z największym krokiem, a jego sąsiadujące węzły z mniejszym. Ostatecznie cała siatka wyjściowa rozwija się, aby reprezentować przestrzeń wejściową.



WYZNACZENIE ZWYCIĘZCY

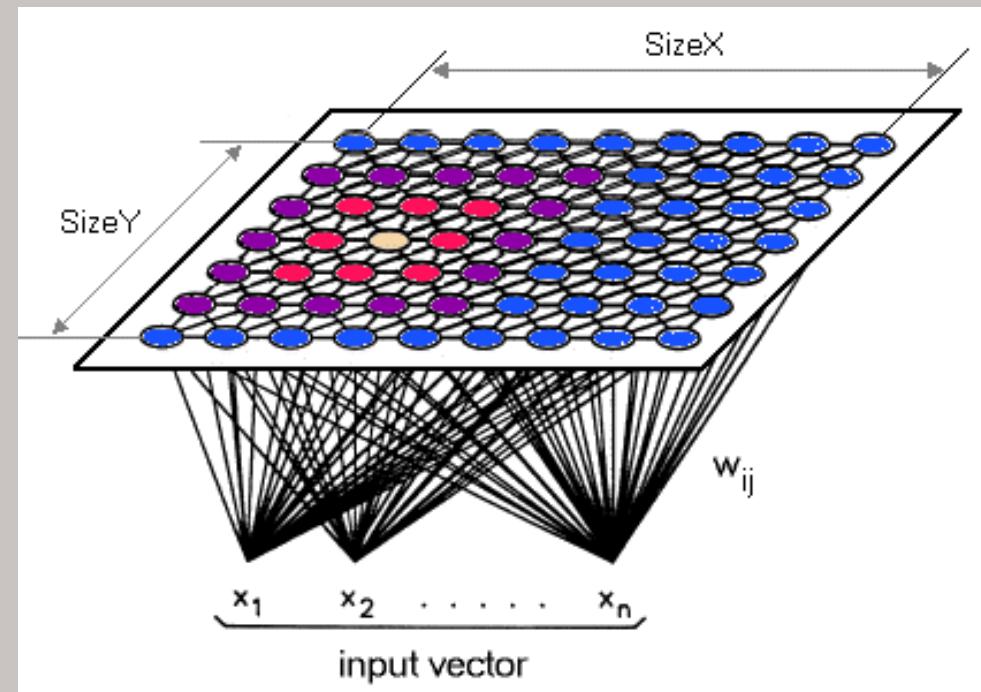
Wyznaczamy kolejno odległości Euklidesa poszczególnych wzorców uczących X_1, \dots, X_k względem wszystkich wektorów wag $W_{1,1}, \dots, W_{I,J}$, gdzie $I \times J$ to ilość węzłów w 2D sieci SOM:

$$d(X_k, W_{i,j}(t)) = \sqrt{\sum_{i=1}^I \sum_{j=1}^J (X_k - W_{i,j}(t))^2}$$

W trakcie obliczania tych odległości wyznaczamy równocześnie argumenty (i, j) dla najbliższego wektora wag (o najkrótszej odległości) do każdego wzorca uczącego:

$$(a, b) = \arg \min_{i,j} d(X_k, W_{i,j}(t))$$

który będzie zwycięzcą aktualizującym swoje wagi najmocniej oraz względem którego będą wyznaczane odległości do sąsiednich węzłów, których wagi będą aktualizowane słabiej.





PARAMETRY ADAPTACJI SIECI SOM

W trakcie uczenia sieci SOM zakres aktualizowanych sąsiadów stopniowo maleje. Na początku zasięg aktualizacji sąsiadów jest duży (może obejmować nawet całą sieć) i stopniowo się zwęża, co można matematycznie wyrazić poprzez zmianę promienia odległości sąsiednich węzłów względem zwycięzcy w zależności od czasu, jaki upłynął od początku nauki t_0 i pewnej stałej zawężania α , np. $\alpha = 1000$:

$\sigma(t) = \sigma_0 \cdot e^{-\frac{t}{\alpha}}$ gdzie σ_0 określa pewien promień początkowy, który na początku może obejmować nawet całą siatkę węzłów, czyli np. $\sigma_0 = \max\{I, J\}$.

Następnie obliczamy współczynnik $\delta(t)$, który zależy od odległości węzła względem zwycięzcy $N_{a,b}(t)$:

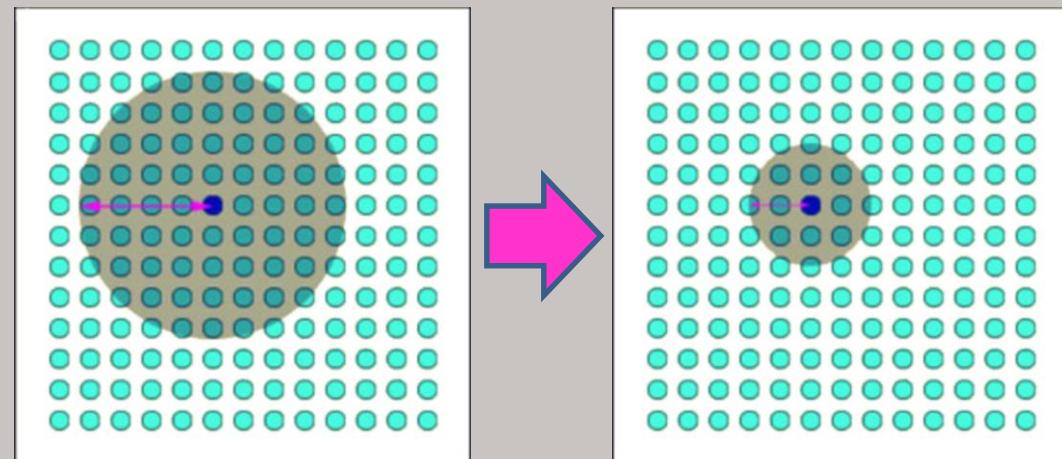
$$\delta(t) = e^{-\frac{d(N_{i,j}(t), N_{a,b}(t))^2}{2 \cdot \sigma^2(t)}}$$

gdzie $N_{i,j}(t)$ to węzeł umieszczony na pozycji (i,j) w 2D siatce (macierzy). Podobnie wyznaczamy współczynnik siły adaptacji wag:

$$\gamma(t) = \gamma_0 \cdot e^{-\frac{t}{\alpha}} \quad \text{np. } \gamma_0 = 1$$

Wagi aktualizujemy w kolejnym dyskretnym kroku czasu $t+1$ według zależności:

$$W_{i,j}(t+1) = W_{i,j}(t) + \delta(t) \cdot \gamma(t) \cdot (X_k - W_{i,j}(t))$$



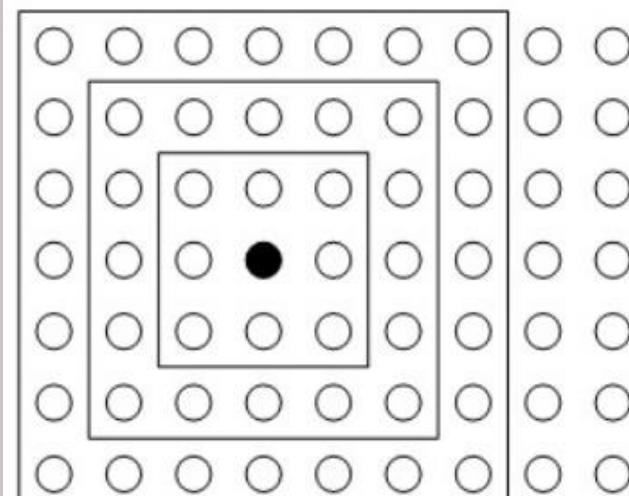


OBLICZANIE ODLEGŁOŚCI

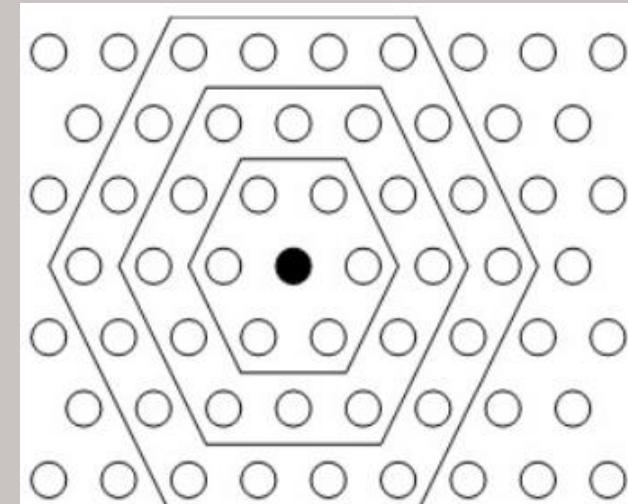
Obliczanie odległości węzłów $N_{i,j}(t)$ względem zwycięzcy $N_{a,b}(t)$ dokonujemy w zależności od przyjętej siatki, oraz miary odległości (Euklidesowej, Manhattan lub innej metryki) np.:

$$d(N_{i,j}(t), N_{a,b}(t)) = |i - a| + |j - b|$$
$$d(N_{i,j}(t), N_{a,b}(t)) = \sqrt{(i - a)^2 + (j - b)^2}$$

Istnieją różne rodzaje siatek węzłów oraz różne sposoby określania najbliższych, sąsiednich:



Siatka prostokątna



Siatka heksagonalna

RODZAJ I WYMIAR SIATKI



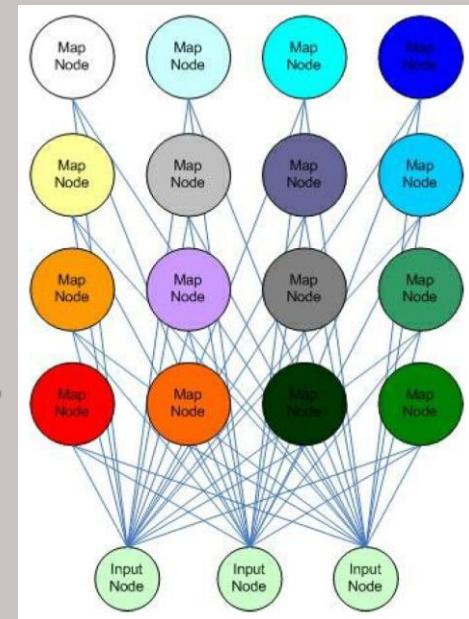
Jak dobrać odpowiednią siatkę węzłów i jej wymiar:

- Warto określić **ilość atrybutów niezależnych m w wektorach/macierzach danych wejściowych**, przy czym $1 \leq m \leq n$.
- Żeby umożliwić sieci SOM nie tylko określenie grup wzorców, lecz również prawidłowe odwzorowanie odległości pomiędzy grupami, należy odpowiednio określić **wymiar przestrzeni siatki v**, na jaką będą rzutowane wzorce wejściowe.

Ten wymiar na pewno nie może być mniejszy niż ilość atrybutów niezależnych, czyli $m \leq v \leq n$. Wobec tego naukę można rozpocząć od wymiaru m, w razie potrzeby stopniowo go zwiększając, lecz nie więcej niż do wymiaru n.

Jak rozpoznać, że wymiar siatki węzłów jest odpowiedni czy nieodpowiedni:

- Rozmiar siatki jest nieodpowiedni, jeśli w wyniku wielu adaptacji rozpoczynających się od różnie wylosowanych wag początkowych odległość węzłów zwycięskich reprezentujących najmocniejsze klasy znacznie się różni. Węzły te nie muszą być w tych samych miejscach, lecz powinny być mniej więcej równo odległe.





JAK DOBRAĆ PARAMETRY?

Z punktu widzenia adaptacji sieci SOM ważny jest dobór odpowiednich parametrów:

- zakres zmienności początkowo wylosowanych niewielkich dodatnich wag różnych od zera, przy czym zakres losowania tych wartości powinien być najlepiej z zakresu od setnych do dziesiętnych zakresu zmienności dla danego atrybutu,
np. jeśli $x_i^{min} \leq x_i \leq x_i^{max}$, wtedy wagi powinny być losowane z zakresu, czyli:

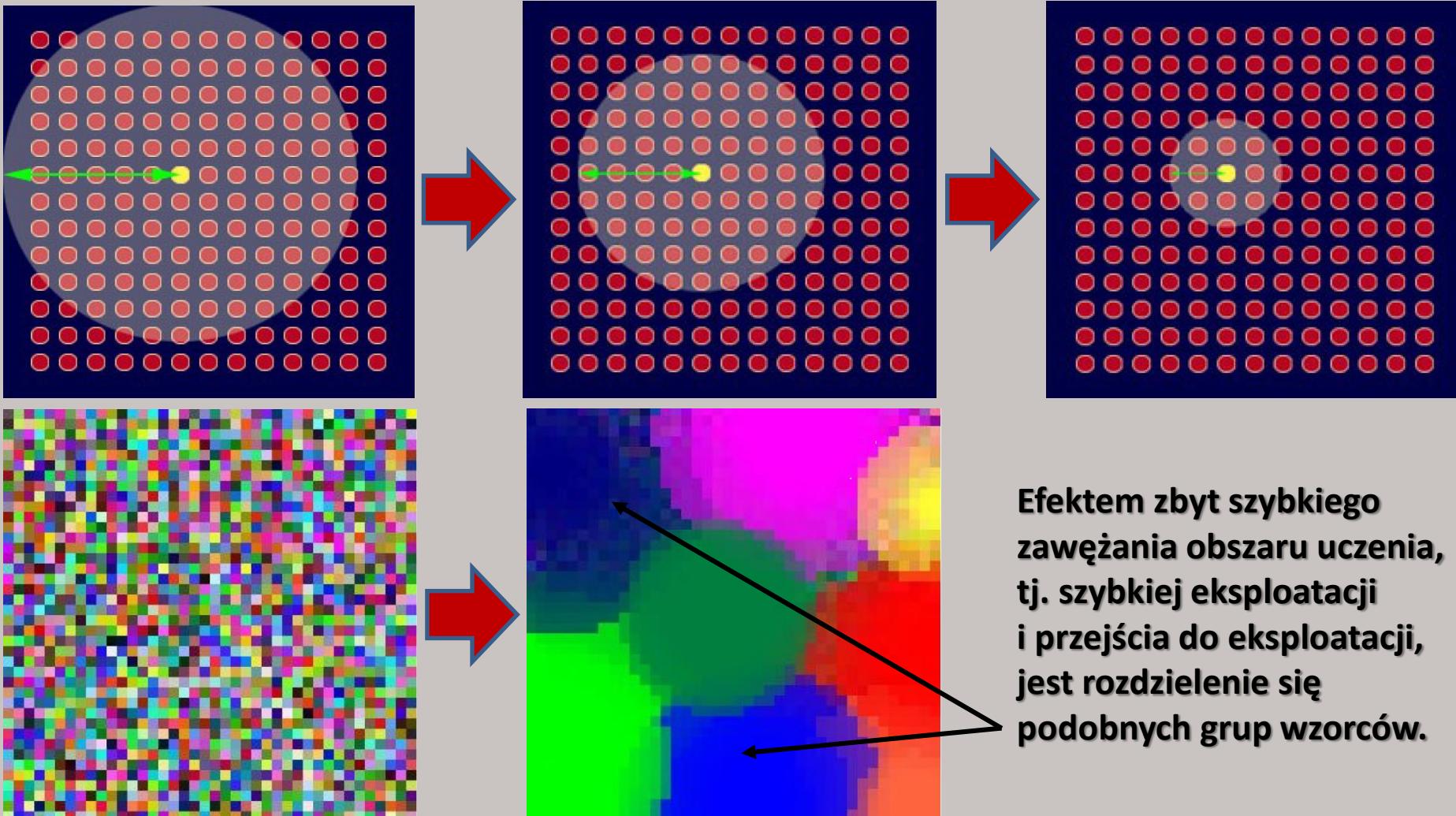
$$\frac{x_i^{max} - x_i^{min}}{100} \leq w_i \leq \frac{x_i^{max} - x_i^{min}}{10}$$

- niezależność atrybutów danych wejściowych $X_k = \{x_1, x_2, \dots, x_n\}$ (gdzie n to rozmiar danych wejściowych),
- szzybkość adaptacji i dobór współczynników, tj.: α , przy czym w trakcie nauki $t \leq \alpha$,
- określenie ilości stopni swobody (rodzaju siatki oraz jej wymiaru), czyli potencjalnie ilości najbliższych sąsiadów o podobnym stopniu bliskości.
- ilości węzłów wyjściowych siatki, których ilość powinna być znaczco mniejsza niż ilość wzorców uczących po to, żeby „zmusić” sieć do reprezentacji podzbiorów podobnych wzorców przy pomocy tych samych węzłów (kompresja reprezentacji). Ilość tych węzłów nie może też być mniejsza niż ilość spodziewanych klas oraz wymiarowość danych.
Jeśli znana jest liczba pożądanych klas/grup c, wtedy początkowa ilość węzłów może być określona jako c^m tworząc m-wymiarową hiper kostkę (macierz) węzłów o wymiarze c, zamiast tworzenia siatki 2D.



EKSPLORACJA >>> EKSPLOATACJA

Najpierw siatka musi pogrupować podobne wzorce, więc obszar uczenia jest szeroki (eksploracja), następnie zawężamy go w celu przyspieszenia nauki i wyspecjalizowania się poszczególnych węzłów sieci (eksploatacja):





WYMIAR I WIELKOŚĆ SIATKI

Jak rozpoznać, że ilość węzłów siatki jest niewystarczająca?

- Jeśli w wyniku nauki dla wielu adaptacji rozpoczynających się od różnie wylosowanych wag początkowych powstają różni lub różnoliczni zwycięzcy w stosunku do innych wyników, oznacza to trudność, którą można rozwiązać stopniowo zwiększając ilość węzłów, lecz nie więcej niż ilość wzorców / 2, a dla bardzo licznych zbiorów wzorców nawet ich pierwiastek, czyli: $\sqrt{\text{ilosc wzorców}}$.

Sposób dobierania ilości węzłów i wymiaru siatki:

- Rozpoczynamy od siatki o wymiarze m równym ilości atrybutów niezależnych, a jeśli jest on nieznany wtedy od wartości 1, 2 lub $\sqrt{\text{ilosc atrybutow}}$.
- Początkową ilość węzłów dla danego wymiaru c dobieramy według ilości spodziewanych/pożądanych klas/grup, a więc dla całej siatki c^m .
- Następnie przechodzimy do procesu nauki, badając i porównując ilości reprezentantów dla poszczególnych węzłów po pewnej ilości kroków, np. 1000, rozpoczynając naukę sieci SOM wielokrotnie (np. 10x) od różnie wylosowanych wag początkowych.
- Jeśli otrzymujemy znacznie różniczącą się ilość reprezentantów dla poszczególnych węzłów wyjściowych, należy zwiększyć ilość węzłów w którymś lub kilku wymiarach np. o 1.



WYMIAR I WIELKOŚĆ SIATKI

5. **Gdy ilość reprezentantów (wzorców) poszczególnych węzłów (czyli ilości wzorców reprezentowanych przez poszczególne węzły się w miarę ustabilizuje, wtedy warto policzyć odległości zwycięzców, jako najmniejszą ilość przejść (krawędzi) oddzielających te węzły. Warto porównywać tych zwycięzców, którzy reprezentują podobne zbiory wzorców wejściowych o podobnej ilości.**
6. **Jeśli te odległości znacznie się różnią, oznacza to, iż wymiar hiperprzestrzeni m do której zrzutowaliśmy zbiór wzorców uczących jest niewystarczający do poprawnego odwzorowania relacji pomiędzy grupami/klasami.**
7. **Należy więc zwiększyć wymiar hiperprzestrzeni na m+1 i równocześnie zmniejszyć ilość węzłów dla wszystkich wymiarów i rozpocząć od punktu 2.**
8. **Gdy odległości pomiędzy zwycięzczami reprezentującymi największą ilość wzorców się ustabilizują dla kilku procesów uczenia rozpoczętych od różnych wag losowych, wtedy najprawdopodobniej udało nam się osiągnąć właściwy wymiar podprzestrzeni siatki SOM oraz prawidłową ilość węzłów.**
9. **Ponadto można eksperymentować jeszcze z określeniem sąsiedztwa w takiej siatce, czy tylko rozważamy sąsiedztwo ortogonalne czy również po przekątnej. Generalnie to po przekątnej powinno dawać lepsze wyniki i nie powodować konieczności zbyt dużego zwiększania wymiaru siatki SOM, przyspieszając naukę.**
Taki proces adaptacji może wymagać kilkudziesięciu a nawet kilkuset adaptacji sieci SOM rozpoczętej od różnych wag w celu określenia poprawnego (optymalnego) modelu.



LITERATURA I BIBLIOGRAFIA

1. Shyam M. Guthikonda, Kohonen Self-Organizing Maps, 2005
2. Mat Buckland, <http://www.ai-junkie.com/ann/som/som1.html>
3. ET-Map – <http://ai.eller.arizona.edu/research/dl/etspace.htm>