**Sztuczne sieci neuronowe** - jedna z technik eksploracji danych służąca głównie do klasyfikacji. Prymitywna próba naśladowania uczenia nieliniowego, występującego w naturze. Sieć nadaje się do szacowania i przewidywania badanych wartości.

Sztuczny neuron zbiera sygnały wejściowe z poprzedzających go neuronów i tworzy z nich jedną wartość za pomocą pewnej funkcji (np. sumy). Wynik jest argumentem tzw. funkcji aktywacji (najczęściej nieliniowa), która zwraca sygnał wyjściowy, przekazywany do następnych neuronów. Sieci neuronowe są odporne na zaszumiane.

Sieć składa się z wielu węzłów, z wagami przypisanymi do każdego połączenia.

**Kodowanie** - sygnały obsługiwane przez sieć neuronową muszą mieścić się w zakresie [0, 1]. Należy zatem dane ciągłe zestandaryzować lub znormalizować (np. min-max). Dla atrybutów jakościowych można zastosować znaczniki (flagi, np. czy mężczyzna, czy kobieta). Niewłaściwe przypisywanie wartości liczbowych do zmiennych jakościowych może prowadzić do wielu błędów.

Neurony wyjściowe zawsze zwracają ciągłą wartość z zakresu [0, 1].

**Sieć warstwowa** - składa się z dwóch lub więcej warstw (najczęściej 3, wej, ukryta, wyj).

**Sieć jednokierunkowa** - przepływ informacji dozwolony jest tylko w jednym kierunku (brak sprzężeń zwrotnych)

**Sieć pełna** - każdy neuron w warstwie jest połączony z każdym neuronem warstwy kolejnej. Brak połączeń między neuronami w danej warstwie.

Każdemu połączeniu przypisana jest waga.

Liczbę warstw ukrytych dobiera się arbitralnie (doświadczalnie). Większa liczba neuronów w warstwie ukrytej zwiększa moc obliczeniową sieci (do skomplikowanych wzorców). Zbyt duża ich liczba prowadzi do przeuczenia. Jeśli klasyfikacja jest niedokładna należy zwiększyć liczbę neuronów w warstwie ukrytej.

**Funkcja aktywacji** - modelowanie pracy rzeczywistego neuronu, w którym przekazanie sygnału następuje dopiero po przekroczeniu przez potencjał wejściowy pewnej wartości progowej (najczęściej sigmoidalna).

**Uczenie sieci** - w sposób nadzorowany, potrzebny jest duży zestaw danych kompletnych ze zmienną celu (zbiór uczący). Każdy rekord uczący jest przepuszczany przez sieć, otrzymywany jest wynik i obliczany błąd.

Uczenie sprowadza się do poszukiwania zbioru wag, które będą minimalizować błąd. Jest to tzw. propagacja wsteczna.

**Reguła największego spadku** - wskazuje kierunek dostosowywania wag, prowadząc do zmniejszenia błędu. Wyznacza się gradient błędu względem wektora wag *w* otrzymując wektor cząstkowych pochodnych błędu względem każdej wagi.

**Współczynnik uczenia** - regulator wielkości przesunięcia wagi. Jest to wartość stała z zakresu [0, 1]. Pomaga przesunąć wagi w kierunku minimum globalnego. Współczynnik mnoży się przez wartość pochodnej cząstkowej.

**Propagacja wsteczna** - na podstawie błędu na wyjściu oblicza się błąd dla poszczególnych neuronów poprzednich warstw od warstwy wyjściowej (błąd przekazywany jest wstecz).

**Składnik momentu** - ma na celu poprawę skuteczności alg. propagacji wstecznej. Sprawia, że krok uczenia jest średnią wykładniczą wszystkich poprzednich kroków uczenia.

**Grupowanie** - jedna z metod eksploracji danych, oznacza grupowanie rekordów, obserwacji lub przypadków w klasy podobnych obiektów. Celem metod grupowania jest identyfikacja grup rekordów aby ich podobieństwo było bardzo duże wewnątrz grupy, a do rekordów z innej grupy bardzo małe. Chodzi o to by stworzyć grupy, dla których zmienność pomiędzy grupami będzie duża w porównaniu do zmienności wewnątrzgrupowej.

**Grupa** - to zbiór rekordów podobnych do siebie nawzajem i niepodobnych do rekordów z innych grup.

**Różnica między klasyfikacja a grupowaniem**

Brak zmiennej celu w przypadku grupowania. Głównym zadaniem algorytmu grupowania jest dokonanie podziału zbioru na stosunkowo zgodne grupy.

**Analiza skupień** - to technika pozwalająca zredukować przestrzeń przeszukiwań dla algorytmów.

Aby przeprowadzić analizę skupień należy wziąć pod uwagę:

1. Jak mierzyć podobieństwo.
2. Jak kodować zmienne jakościowe.
3. Jak normalizować zmienne ilościowe.
4. Ile grup należy się spodziewać

**Podział algorytmów grupowania**

* hierarchiczne
* niehierarchiczne

**Grupowanie hierarchiczne** - polega na tworzeniu drzewiastej struktury poprzez rekurencje:

* łączenie (**metody aglomeracyjne**)
* dzielenie (**metody rozdzielające**)

istniejących grup.

**Metoda aglomeracyjna** - (ześrodkowanie) zakłada, że początkowo każda obserwacja to osobna grupa składająca się z jednego elementu. W kolejnych krokach dwie grupy będące najbliżej siebie łączy się w nową wspólną grupę. W każdym kroku uzyskuje się redukcję (o jeden) liczby skupień (grup) w zbiorze danych.

**Metody rozdzielające** - cały zbiór umieszczany jest w jednej grupie. W kolejnych krokach rozdziela się rekordy najbardziej niepodobne do siebie na oddzielne dwie grupy. Algorytm kończy działanie do momentu kiedy powstaną grupy jedno elementowe.

**Kluczowa rola to mechanizm określania odległości między grupami rekordów.**

**Określanie odległości**

**Metod pojedynczego połączenia** - metoda najbliższego sąsiedztwa. Określa odległość pomiędzy najbliższymi punktami. Oparta na minimalnej odległości pomiędzy dowolnym rekordem grupy A i dowolnym rekordem z gr. B. (Tendencja do tworzenia długich cienkich grup).

**Metoda całkowitego połączenia** - metoda najdalszego sąsiedztwa. Określa ona odległości pomiędzy najbardziej oddalonymi punktami. Oparta na maksymalnej odległości pomiędzy dowolnym rekordem z grupy A i dowolnym rekordem z gr. B. (Tendencja do tworzenia zwięzłych kulistych grup).

**Metoda średniego połączenia** - oparta na wyliczaniu średniej odległości wszystkich rekordów z grupy A do wszystkich rekordów z gr. B. Metoda pozwala na ograniczenie wpływu rekordów o wartościach ekstremalnych na końcowy wynik grupowania.

**Algorytm *k*-średnich**

1. Ustalenie na ile grup (*k*) należy podzielić zbiór danych.
2. Losowy wybór *k* rekordów z przeznaczeniem na środki grup.
3. Znalezienie najbliższego środka grupy dla każdego rekordu (Najczęściej odległość euklidesowa).
4. Znalezienie (dla każdej z *k* grup) centroidu grupy i uaktualnienie położenia każdego środka grupy tak, aby stała się nim wartość znalezionego centroidu.
5. Powtarzanie kroków 3 i 4 do zbieżności lub planowanego zakończenia algorytmu.

**Centorid** - środek ciężkości wszystkich punktów w grupie.

Algorytm kończy działanie wtedy kiedy centroidy przestaną się zmieniać lub wtedy kiedy zostanie osiągnięte kryterium zbieżności (np. brak istotnego zmniejszenia sumarycznego błędu kwadratowego).

**Zagrożenia związane ze stosowaniem algorytmu *k*-średnich**:

1. Nie gwarantuje znalezienia minimum globalnego SSE, może zatrzymać się na minimum lokalnym. Aby tego uniknąć należy uruchomić algorytm kilkukrotnie z różnymi początkowymi środkami grup.
2. Liczbę grup ustala się arbitralnie, co może prowadzić do nieprawidłowego podziału danych. W razie gdy nie zna się liczby występujących grup, należy dokonać podziału dla różnych wartości *k* i wybrać ten wynik, dla którego SSE jest najmniejsze.
3. Zależnie od sformułowanego problemu pewne atrybuty mogą być ważniejsze od innych. W takiej sytuacji, ponieważ przynależność do grupy zależy od odległości od jej środka można wprowadzić wagi (stosować różne metody rozciągania osi).

**Sieci samoorganizujące** - specyficzne klasy sztucznych sieci neuronowych tzw. sieci samoorganizujące się. Celem sieci jest przekształcanie złożonych, wielowymiarowych sygnałów wejściowych w prostsze, mniej wymiarowe dyskretne odwzorowania. Organizują neurony wyjściowe w grupy, w których neurony położone bliżej siebie są bardziej podobne do siebie nawzajem niż do innych neuronów, położonych dalej.

Sieć opiera działanie na uczeniu z rywalizacją, gdzie neurony wyjściowe rywalizują między sobą, tak aby zostać neuronem wygrywającym (będzie on jedynym węzłem pobudzanych przy konkretnej konfiguracji zmiennych wejściowych).

Sieć jest jedno kierunkowa (nie zawiera sprzężenia zwrotnego) i pełna (każdy neuron jest połączony ze wszystkimi neuronami warstwy następnej). Każde połączenie ma wagę.

Wartości zmiennych muszą być normalizowane lub standaryzowane.

Sieci samoorg. nie podsiadają warstwy ukrytej.

**Główne procesy**

1. **Rywalizacja** - neurony wyjściowe rywalizują ze sobą, aby uzyskać najlepszą wartość danej funkcji decyzyjnej (odległość euklidesowa, zwycięża najmniejsza wartość).
2. **Współdziałanie** - neuron wygrywający zostaje środkiem sąsiedztwa pobudzonych neuronów. W sieciach samoorg. wszystkie neurony z sąsiedztwa dzielą się „pobudzeniem” („nagrodą”) neuronu wygrywającego - podlegają adaptacji. Pomimo braku bezpośrednich połączeń neurony warstwy wyjściowej dążą do dzielenia wspólnych cech (na zasadzie „sąsiedztwa”).
3. **Adaptacja** - węzły z otoczenia węzła wygrywającego podlegają uczeniu. Proces ten polega na dopasowywaniu ich wag tak, aby uzyskać poprawę wartości funkcji decyzyjnej.

**Sieci Kohonena** - rodzaj sieci samoorganizującej się. Uczenie odbywa się za pomocą algorytmu Kohonena.

Alg. Kohonena - węzły z otoczenia neuronu wygrywającego są dopasowywane wg zależności:

*i* - numer węzła neuronu wejściowego

*j* - nr węzła neurony wyjściowego

*xi* - wartość *i*-tego pola rekordu danych,

*wij* - waga połączenia pomiędzy *i*-tym neuronem wejściowym, a *j*-tym neuronem wyjściowym,

*η* - współczynnik uczenia; 0 < *η* < 1

W alg. Kohonena współczynnik uczenia jest funkcją malejącą, zależną od etapów uczenia.

**Kroki alg Kohonena**

1. **Rywalizacja** - dla każdego węzła wyjściowego *j* obliczyć wartość funkcji decyzyjnej *D*(*wj*, *x*). Znaleźć neuron wygrywający *J*, który charakteryzuje się najlepszą wartością *D*(*wj*, *x*).
2. **Współdziałanie** - zidentyfikować wszystkie neurony wyjściowe *j* z otoczenia węzła *J*, określone przez rozmiar sąsiedztwa *R*. Dla tych węzłów wykonać należy działania:

**Adaptacja** - dopasowanie wag według wzoru: ↑.

1. Jeśli zachodzi potrzeba, dopasować współczynnik uczenia *η* i rozmiar sąsiedztwa.
2. Zatrzymać jeśli spełnione są warunki „stopu”.