W trakcie zajęć zajmować będziemy się metodami klasteryzacji (ang. *clustering*) i klasyfikacji (ang. *classification*) obiektów. We wszystkich przypadkach analizowane obiekty będą opisane za pomocą wektorów cech w przestrzeni euklidesowej Rn.

* Celem klasteryzacji będzie wykrycie grup obiektów tak, aby podobne do siebie obiekty były w jednej grupie, różne obiekty w różnych grupach. We wszystkich przypadkach podobieństwo będzie mierzone za pomocą metryki euklidesowej w przestrzeni cech Rn. Określenie przynależności obiektów do grup będzie się odbywało przez przypisanie im etykiet w ten sposób, aby obiekty z tej samej grupy miały tę samą etykietę.
* Celem klasyfikacji będzie stworzenie funkcji, która będzie w stanie przypisać etykietę do każdego innego obiektu opisanego w przestrzeni cech Rn. Funkcja ta zwykle tworzona jest automatycznie w oparciu o dane treningowe, to jest dane zawierające przykłady obiektów, które mają wcześniej przypisane, znane etykiety.

W przypadku gdy obiekty opisane są za pomocą cech w przestrzeni dwuwymiarowej (wówczas n=2) możemy je przedstawić w postaci dwuwymiarowego wykresu, gdzie każdemu obiektowi odpowiada jeden punkt. Jeśli obiekt posiada przypisaną, prawdziwą etykietę to możemy te punkty pokolorować tak, aby obiekty o tej samej etykiecie miały ten sam kolor. Dodatkowo jeśli dane poddaliśmy klasteryzacji lub klasyfikacji możemy zwizualizować wyniki tych metod w następujący sposób:

* W przypadku klasteryzacji można posłużyć się diagramem Woronoja (ang. *Voronoi diagram*). Na diagramie tym dla każdego obiektu pokazywane jest jego komórka Woronoja, to jest obszar przestrzeni Rn, w której dla każdego punktu z tego obszaru najbliżej jest do rozważanego obiektu. Jeśli w wyniku grupowania każdy obiekt uzyska etykietę, to można tę informację wykorzystać do pokolorowania komórek tak, aby komórki obiektów z tą samą etykietą miały taki sam kolor:  
  Obraz zawierający mapa, zrzut ekranu, diagram

  Opis wygenerowany automatycznie  
  Jeśli dysponujemy prawdziwymi etykietami obiektów, to one również mogą zostać zaprezentowane poprzez odpowiednie pokolorowanie punktów odpowiadających tym obiektom:  
  Obraz zawierający diagram, tekst, zrzut ekranu, mapa

  Opis wygenerowany automatycznie
* W przypadku klasyfikacji zwykle pokazuje się tak zwaną granicę decyzyjną (ang. *decision boundary*). Ponieważ funkcja klasyfikująca potrafi sklasyfikować każdy punkt przestrzeni Rn, to każdemu punktowi na wykresie można przypisać kolor odpowiadający jego etykiecie. Zbiór punktów na granicach pomiędzy obszarami o tym samym kolorze odpowiada granicy decyzyjnejObraz zawierający tekst, mapa, zrzut ekranu

  Opis wygenerowany automatycznie

Zadanie rozgrzewkowe polega na przygotowaniu funkcji, które będą potrafiły wczytać dane z pliku CSV, a następnie zwizualizować (w przypadku n=2) wyniki działania metody klasteryzacji i metody klasyfikacji. W tym celu trzeba uzupełnić kod skryptu Pythona dołączonego do tego zadania. Zawiera on instrukcje jak powinny działać wspomniane funkcje i pozwala je przetestować. Realizacja tego zadania może być oparta o dowolnie wybrane biblioteki Pythona (na przykład Matplotlib).

Zadanie to nie jest oceniane. Dobrze jest je jednak wykonać, aby przygotować się do realizacji projektów, w których te funkcje będą potrzebne. Dodatkowo, w celu uruchomienia załączonego skryptu, konieczne będzie również przygotowanie sobie środowiska uruchomieniowego Pythona zawierającego odpowiednie biblioteki (NumPy, scikit-learn i inne wykorzystane do rozwiązania zadania), które również przyda się kolejnych projektach.

Jako wynik prac, można indywidualnie wrzucić uzyskane obrazki wynikowe (obrazy PNG) jako rozwiązanie niniejszego zadania.