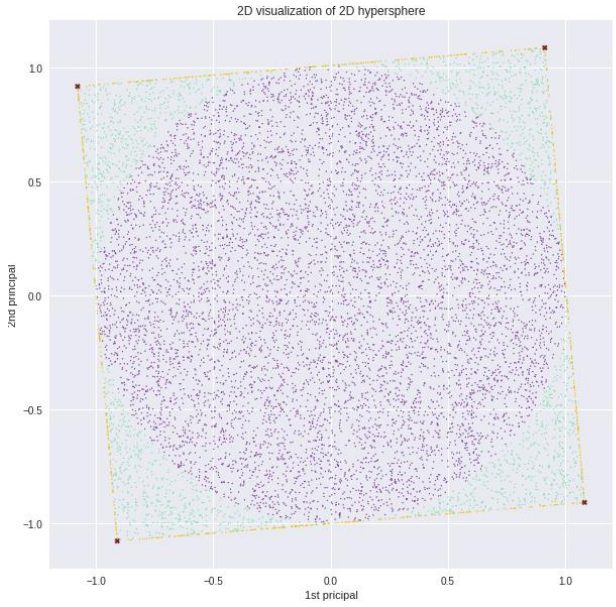
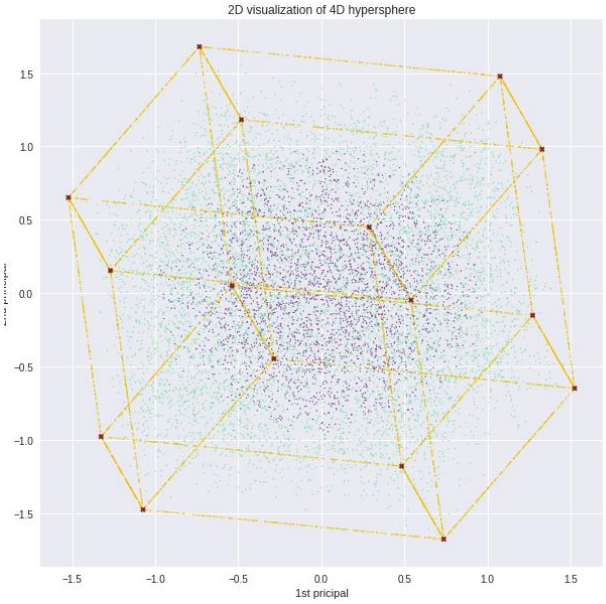
Raport – Condensed Nearest Neighbours
Podstawy nauczania maszynowego
Wyk. Mateusz Woś

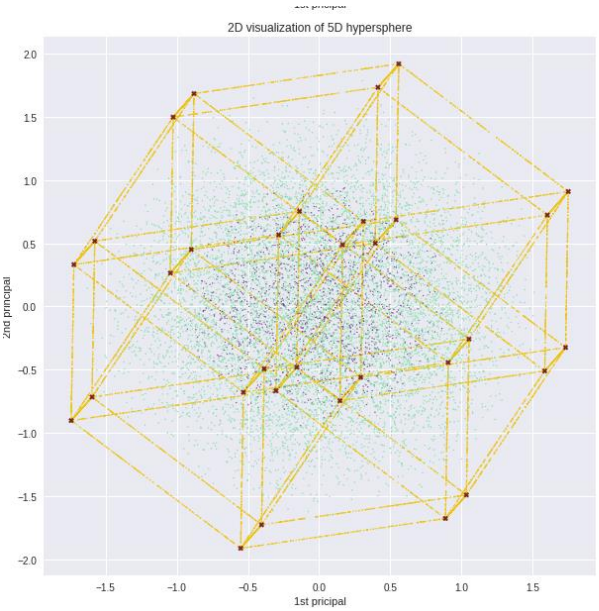
**Zadanie a)**
Wygenerowane dane składają się z 10000 punktów w środku hiperześcianu i dodatkowych 100 punktów na pojedyncza krawędź.
Data colors:{ "corners" -> "dark red", "edges" -> "yellow", "inside" -> "purple", "outside" -> "green"}



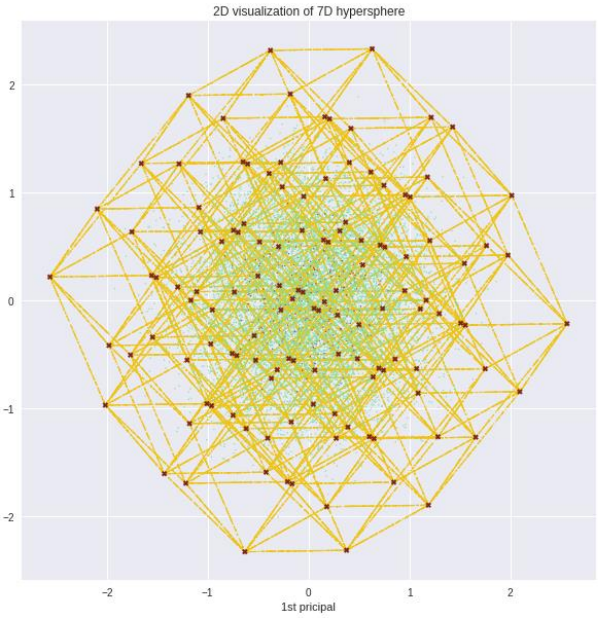

*2D -> 2D*



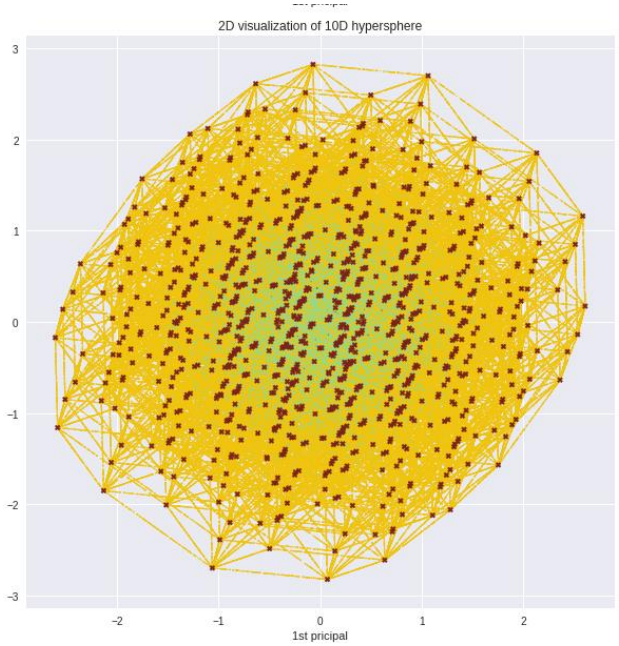

*4D -> 2D*




*5D -> 2D*

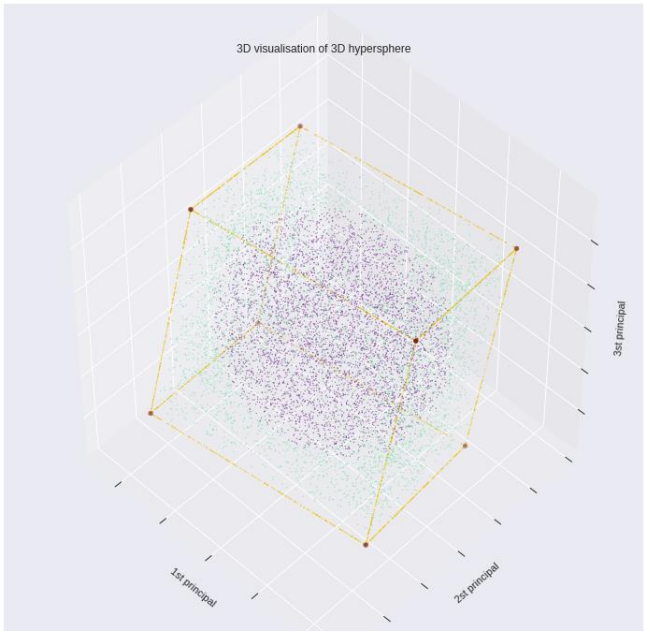



*7D -> 2D*

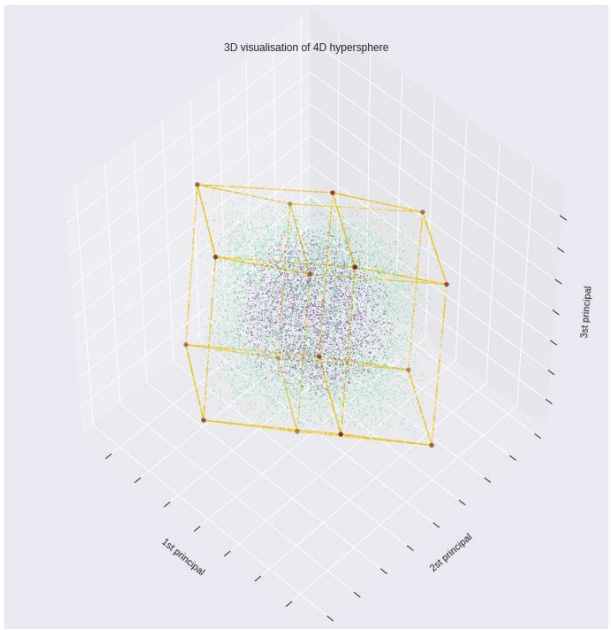



10D -> 2D

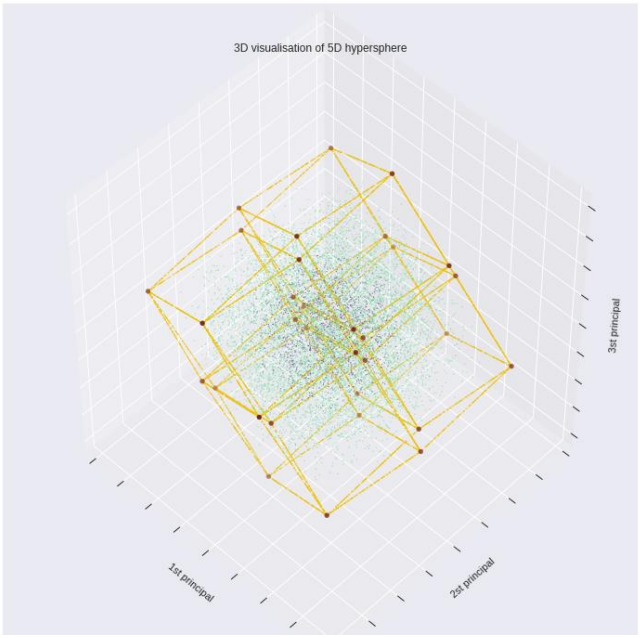
Niestety ze względu na niesamowicie długie czasy transformacji datasetu przez PCA z 13D do 2D byłem zmuszony na delikatna zmianę zadania i wizualizacje rzutu 10D hipersześcianu na 2D płaszczyznę.
Tak samo postąpiłem w wizualizacji rzutów dla 3D – zastąpiłem ostatni podpunkt wersja dla 10D.



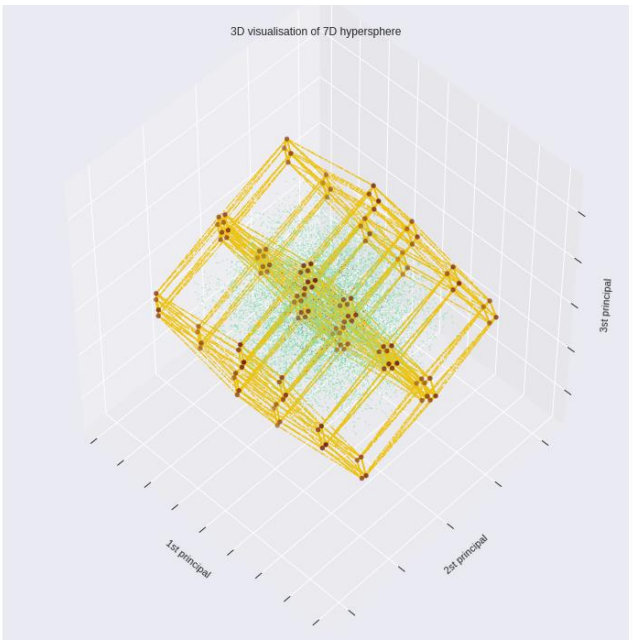

*3D -> 3D*



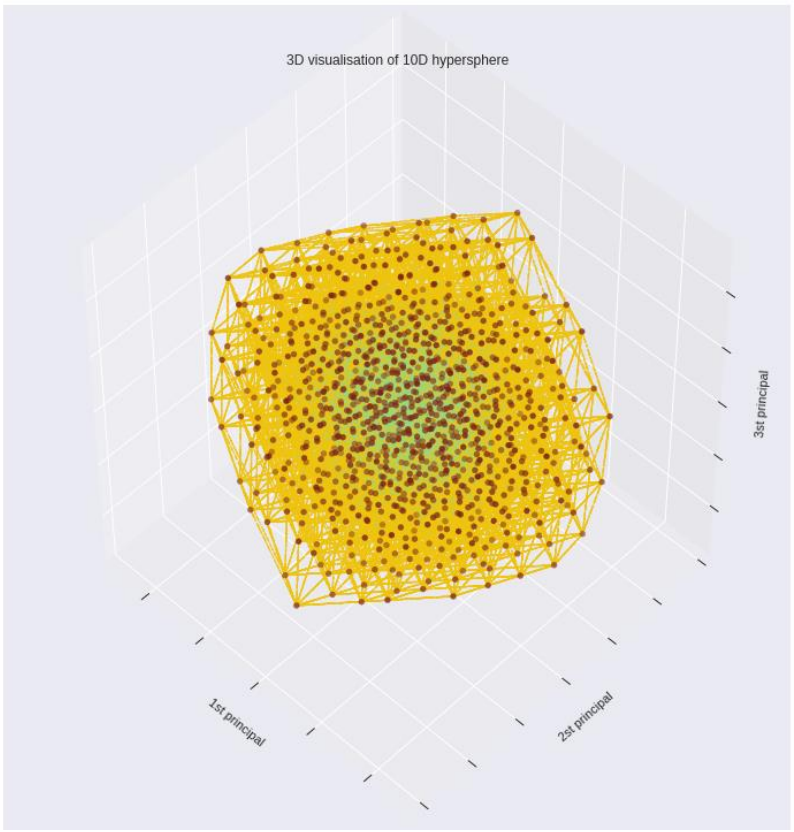

*4D -> 3D*




*5D -> 3D*

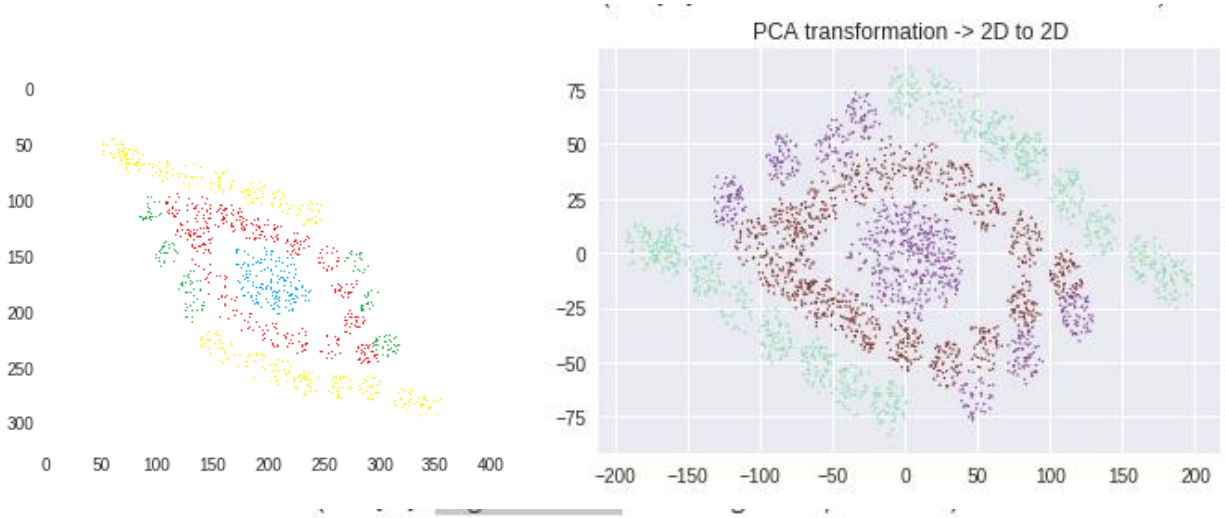


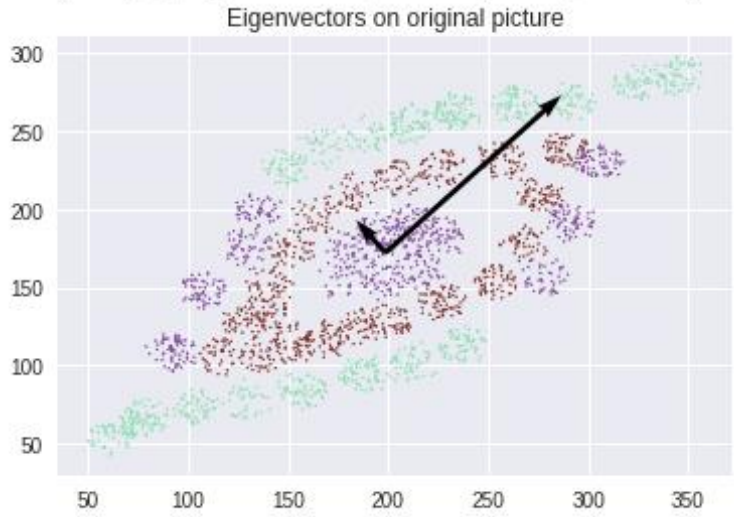

*7D -> 3D*



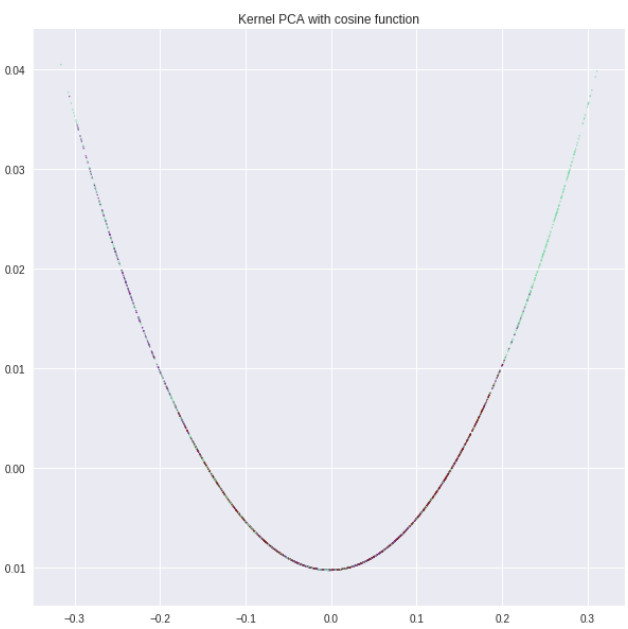

10D -> 3D

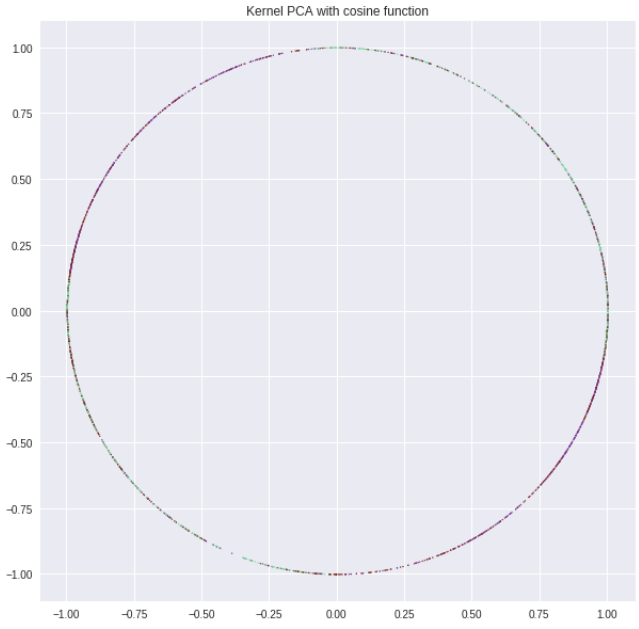
**Zadanie b)**



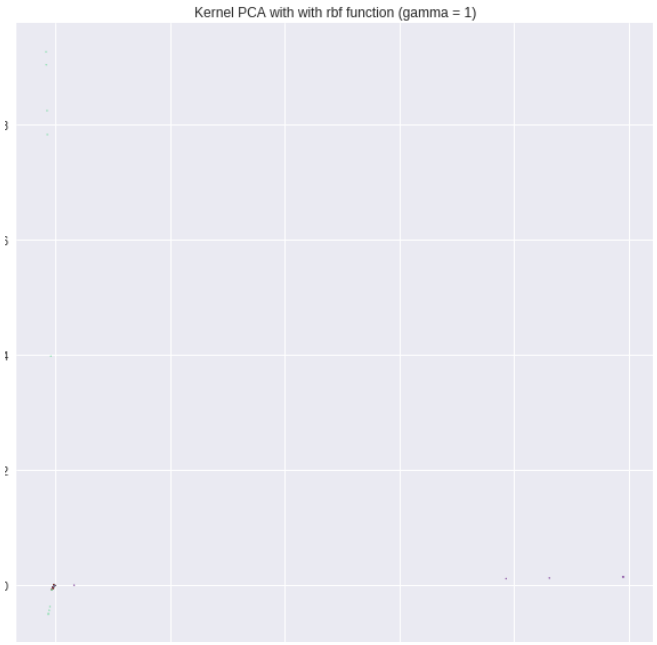



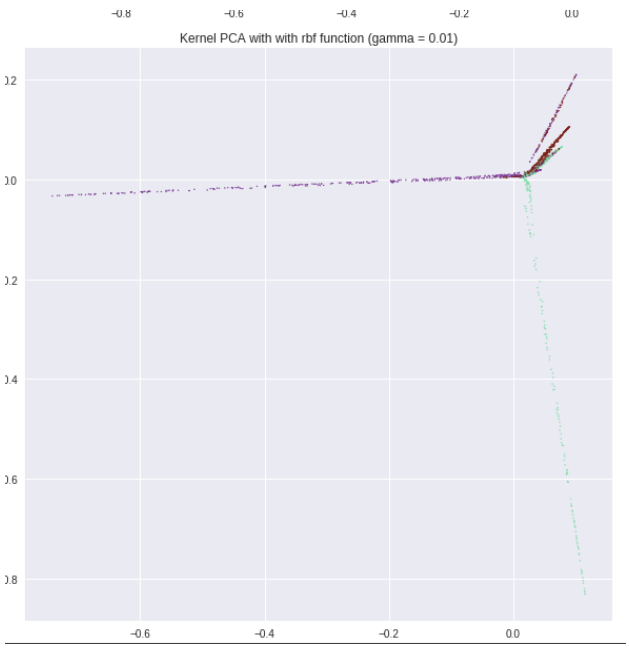

Po przejściu z 2D do 2D widać dokładnie, ze dane są scentralizowane wokół punktu (0, 0)

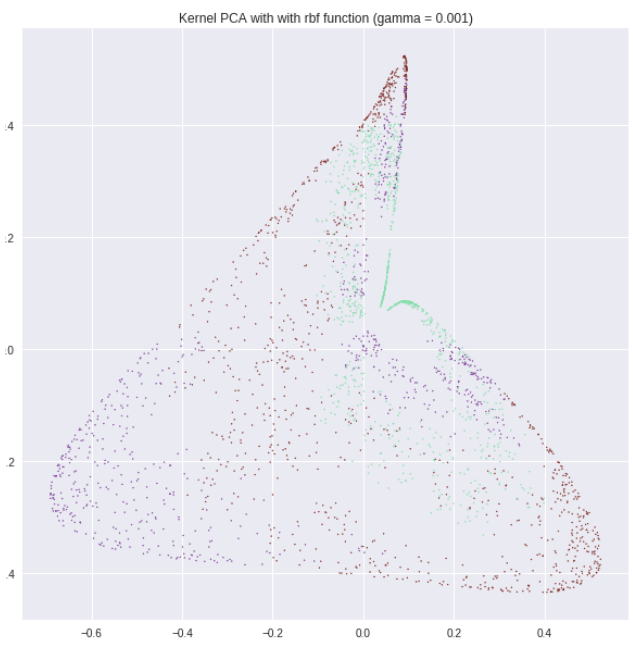


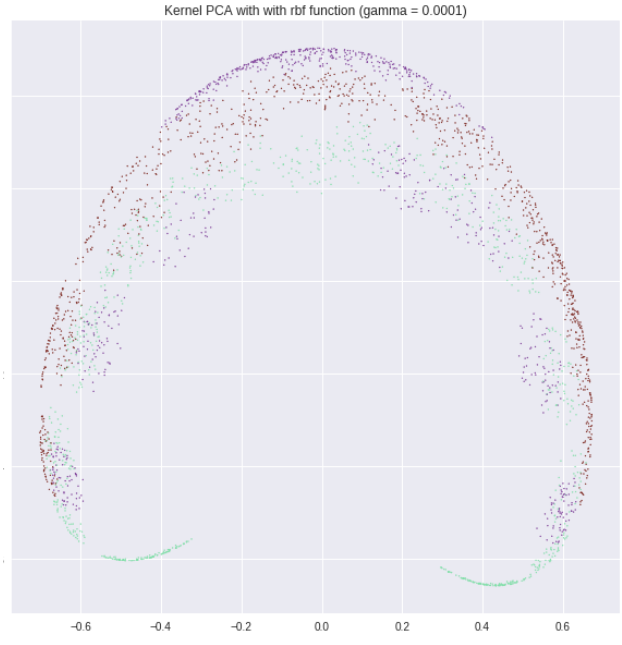



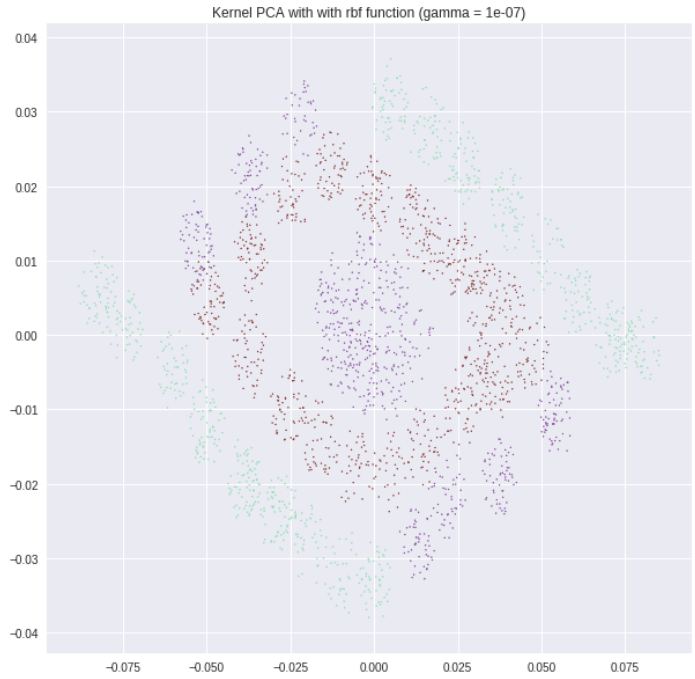

Powyżej widać Cosine karnel PCA przed wyśrodkowaniem danych i po.






Powyżej widać Radial Basis Function karnel PCA z rożnymi parametrami gamma

**Wnioski:**

- Jak można zauważyć na wykresach (jak i zadania z 1 cwiczen) ukazujacych rzut hiperszescianow na plazsczyzny 2 i 3 wymiarowe prawdopodobienstwa pojawienia sie punktow w srodku hiperkuli maleje znaczaca ze wzrostem wymiarow.
- Wraz ze wzrostem wymiarow rzuty stawaly sie coraz bardziej nieczytalene. Rzut stawał sie praktycznie zbiorem samych krawedzi.
- Ograniczenie sie do 2/3 glownych skladowych pozwala nam na usuniecie mniej znaczacych skladowych, dzieki czemu pozbywamy sie cech ktore moglyby naruszyc poprawnosc dzialania naszego modelu, a takze dzieki zmniejszeniu ilosci wymiaru samo przetwarzanie byloby szybsze.

- Przy zastosowaniu PCA z 2D do 2D widzimy, ze baza zostala zmieniona. Obraz zostal skoncetrowany wokol punktu (0, 0)
- Przy użyciu PCA z kernel cosine mozna zauwazyc, iz rozstawienie punktow jest zalezne od kąta pod ktorym one leza od punktu (0, 0) w pierwotnym datasecie. PCA z kernel cosine dla datasetu skoncetrowanego w (0, 0) wyglada jak okrąg, przed wysrodkowaniem jak hiperbola.
- PCA z kernel RBF dla bardzo małych wartości gamma działa podobnie do PCA.
- PCA z kernel RBF dla współczynnika gamma = 1 dla tych danych na wykresie widać bardzo niewiele, wraz z maleniem współczynnika widać separację kolorów. Czym wiekszy wspolczynnik tym slabsza separacja.