Podstawy Uczenia Maszynowego

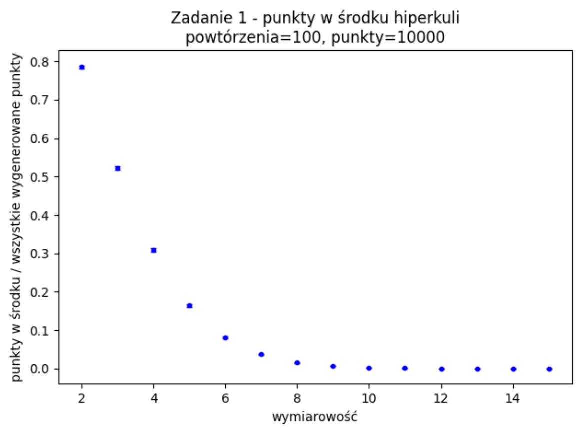
Laboratorium 1 – raport z zadania domowego

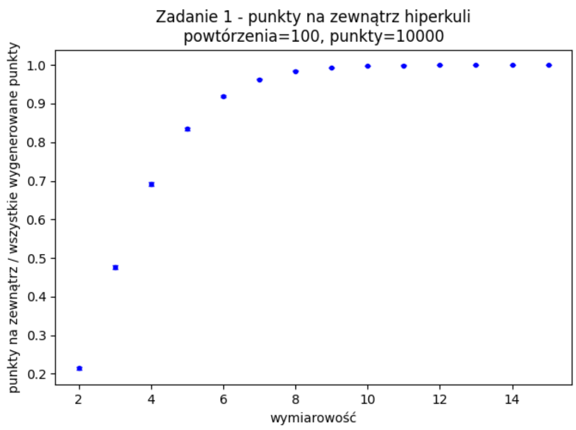
Mateusz Kocot

Spis Treści

Zadanie domowe zaimplementowałem w języku Python. Do stworzenia wykresów wykorzystałem pakiet Matplotlib.

Zadanie 1

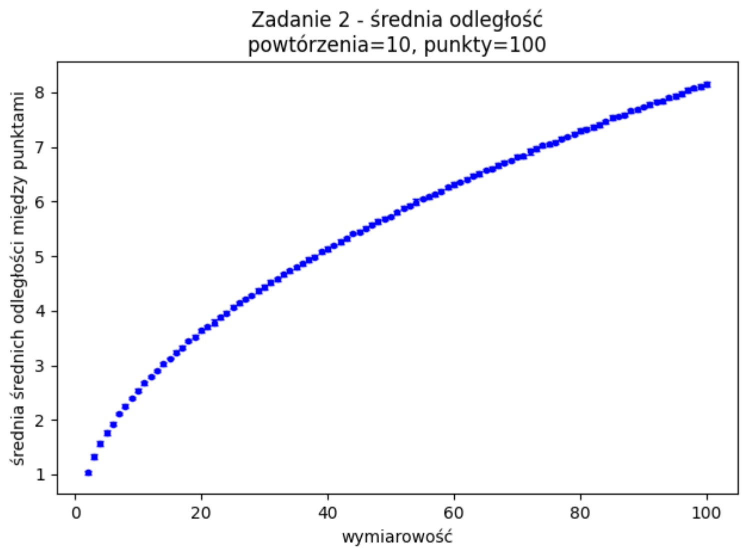




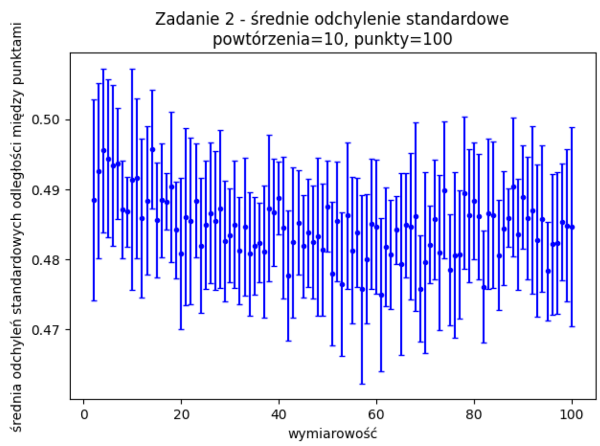
Wraz ze wzrostem wymiarowości maleje liczba punktów znajdujących się w środku hiperkuli. Wzrasta liczba punktów znajdujących się w „narożnikach” hipersześcianu. Od 10 wymiarów, prawdopodobieństwo wylosowania punktu w środku hiperkuli jest już bliskie zeru.

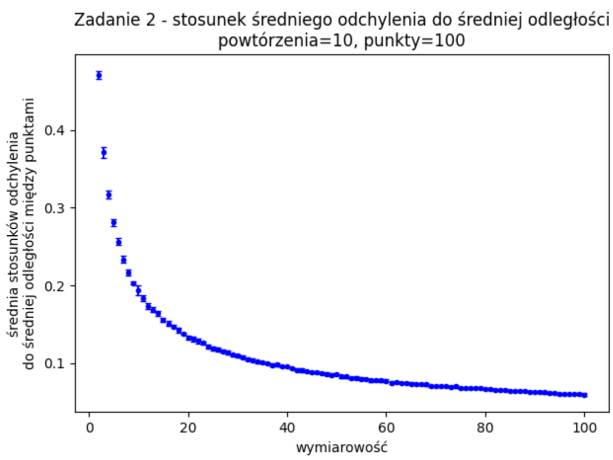
Oznacza to, że przy zwiększaniu liczby wymiarów, hipersześcian staje się objętościowo coraz większy względem hiperkuli.

Zadanie 2



W miarę zwiększania wymiarowości rosną odległości między losowo wybranymi punktami. Każdy kolejny wymiar sprawia, że w trakcie obliczania odległości euklidesowej, suma pod pierwiastkiem będzie miała o jeden składnik więcej, w związku z czym cała odległość także będzie większa.



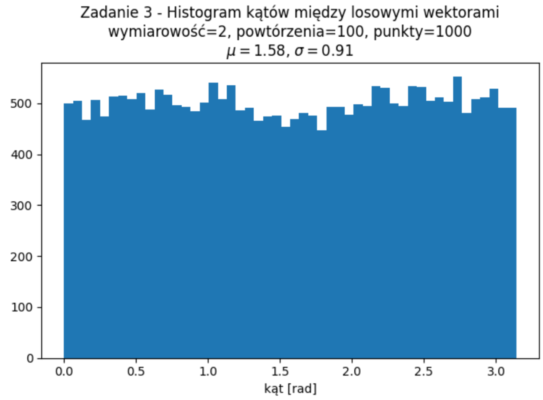


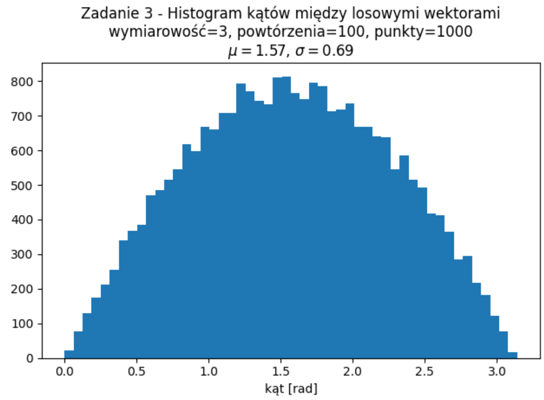
Średnie odchylenie standardowe wydaje się mieć podobną wartość niezależnie od wymiarowości (ok. 0.485).

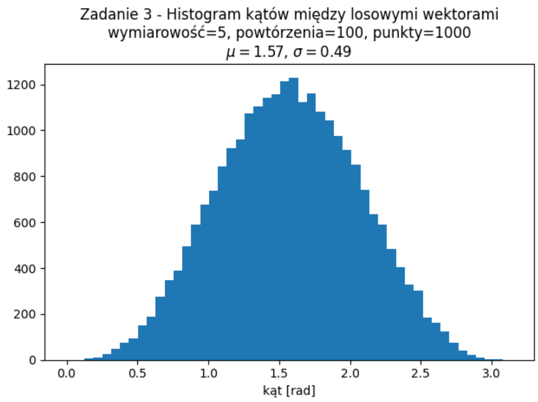
Wraz ze wzrostem wymiarowości rosną odległości, ale odchylenie standardowe pozostaje takie same, więc stosunek średniego odchylenia do średniej odległości maleje. Oznacza to, że różnice w odległościach stają się coraz mniej zauważalne. W konsekwencji, dla dużych wymiarów, odległości między większością punktów w metryce euklidesowej będą bardzo podobne. Własność ta np. utrudnia działanie algorytmom uczenia maszynowego szukającym skupisk.

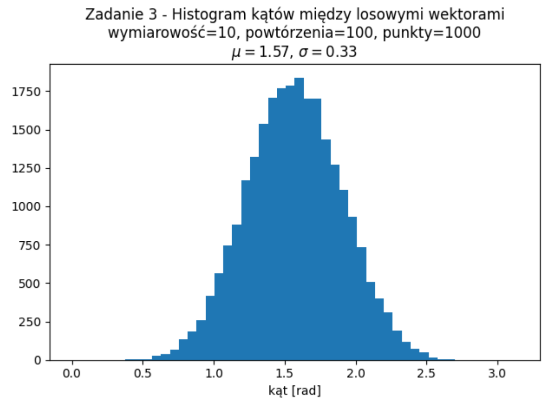
Zadanie 3

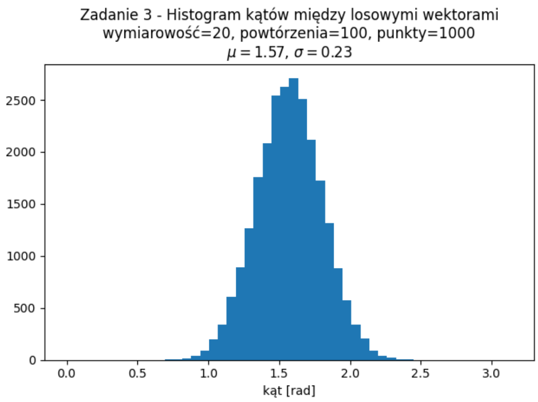
Eksperyment przeprowadzono dla liczby wymiarów z zakresu od 2 do 30. Poniżej przedstawię wybrane histogramy.

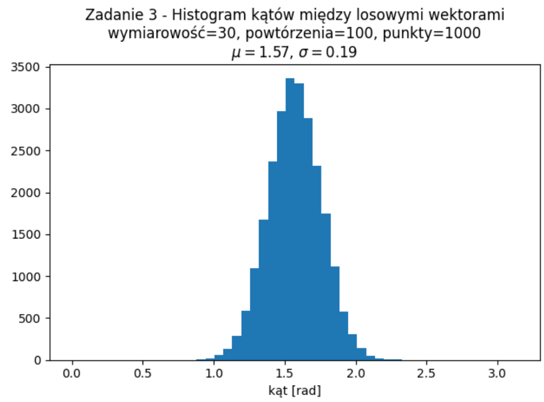












W dwóch wymiarach rozkład otrzymanych kątów jest bliski rozkładowi jednostajnemu o średniej wynoszącej $\frac{\pi}{2} \approx 1.57$. Jednakże, w miarę gdy zwiększana jest wymiarowość przestrzeni, maleje odchylenie standardowe zebranych danych, a to oznacza, że coraz więcej losowo wybranych wektorów jest do siebie prostopadłych.