Machine Learning
lab 1

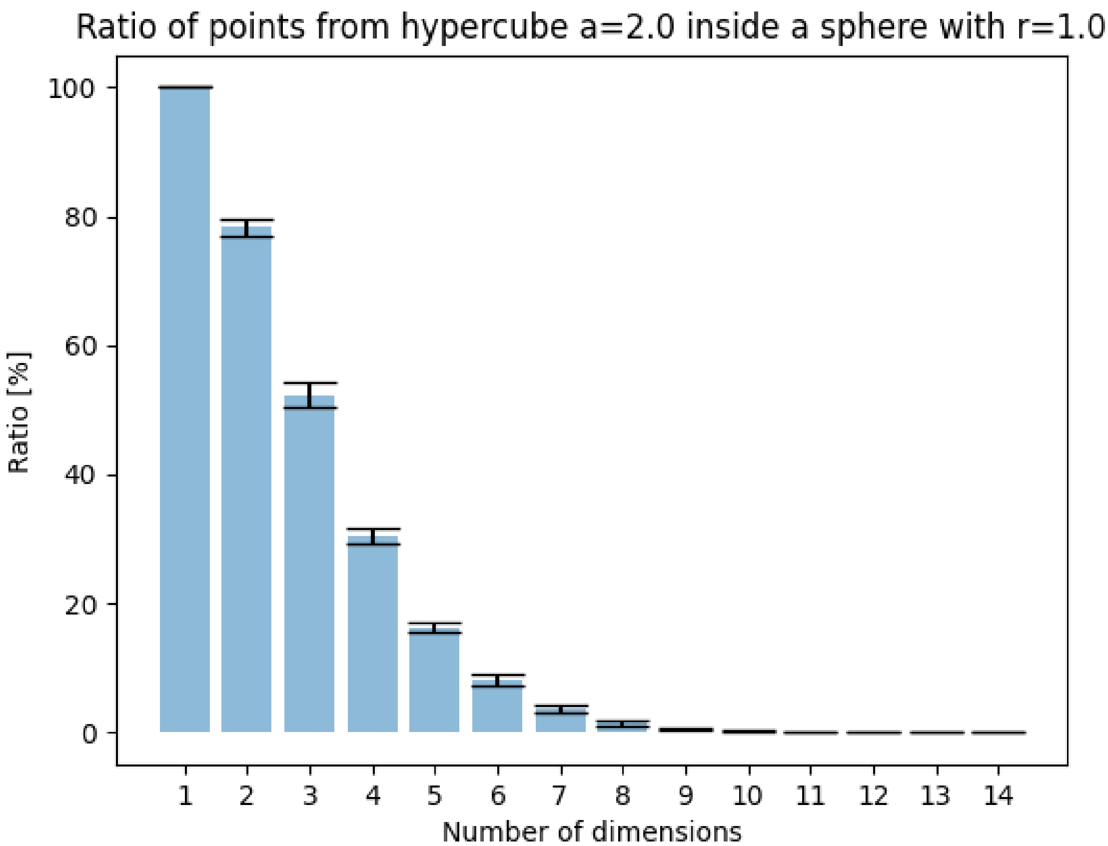
25.03.2021 Mikołaj Zatorski

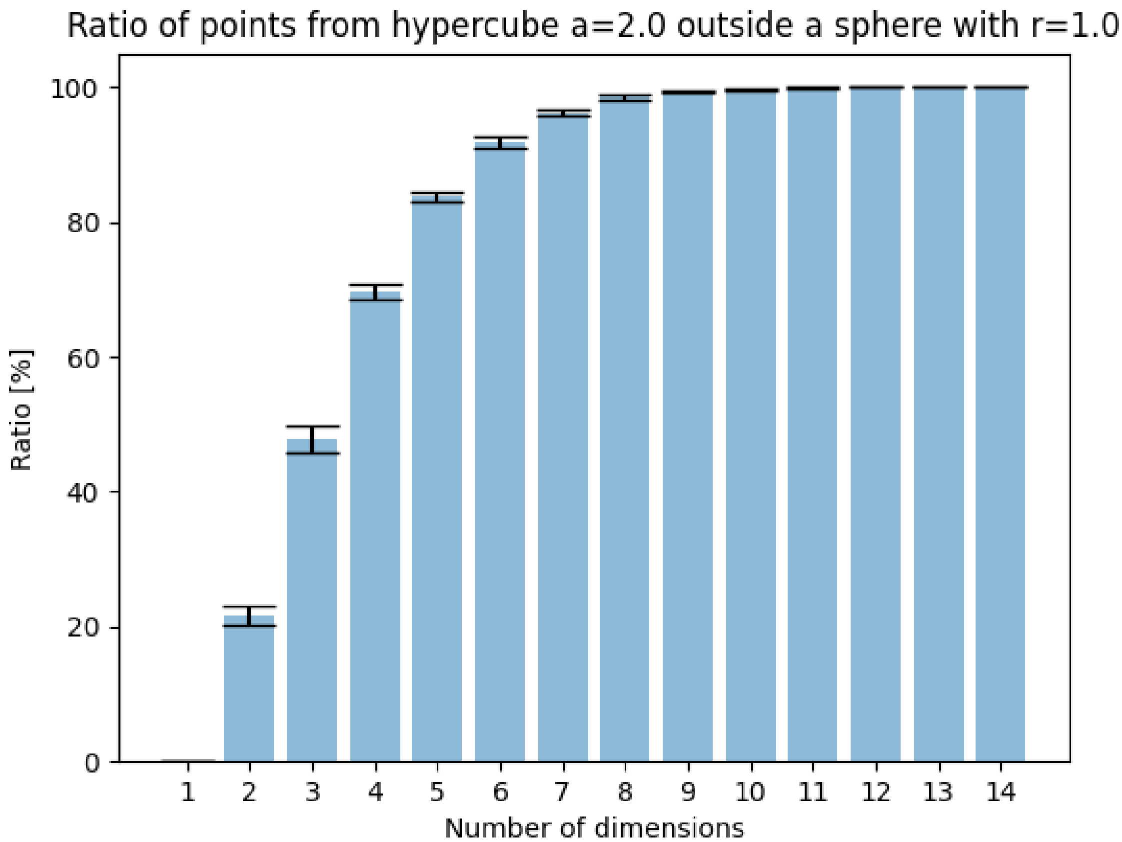
Wszystkie pomiary wykonano dla:

- liczby wymiarów *n*: {1, 2, 3, …, 14}
- liczby wylosowanych punktów: 1000
- liczby powtórzeń losowań i pomiarów: 10

Zad 1.

Wykresy przedstawiające stosunek liczby punktów zawierających się w hiperkuli do liczby punktów wylosowanych z hipersześcianu, w zależności od liczby wymiarów *n.*

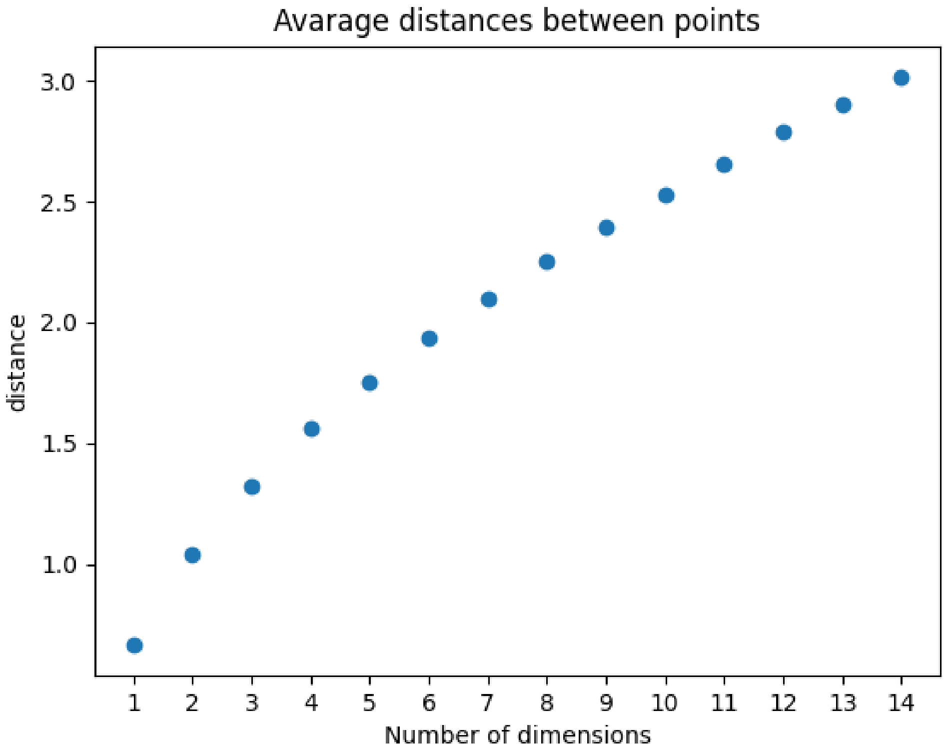




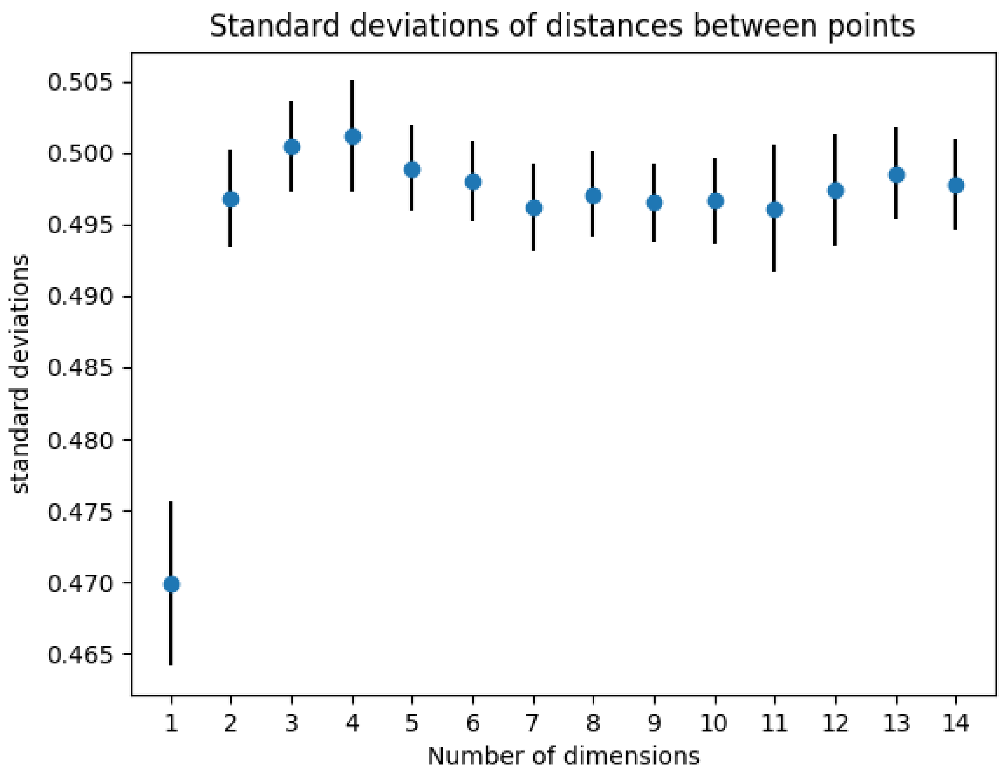
Komentarz:
Jak widać na powyższych wykresach wraz ze wzrostem liczby wymiarów, % punktów zawierających się w hiperkuli maleje. Dla jednego wymiaru wynosi on 100%, dla trzech wymiarów jest to już około 52%, natomiast dla dziewiątego wymiaru i wyższych stosunek ten wynosi praktycznie 0%. Można z tego wywnioskować, że im większy wymiar naszej przestrzeni, tym więcej punktów gromadzi się w "narożnikach" sześcianu, a dla wymiarów powyżej dziesiątego praktycznie wszystkie punktu są osadzone właśnie w "narożnikach".

Zad 2.

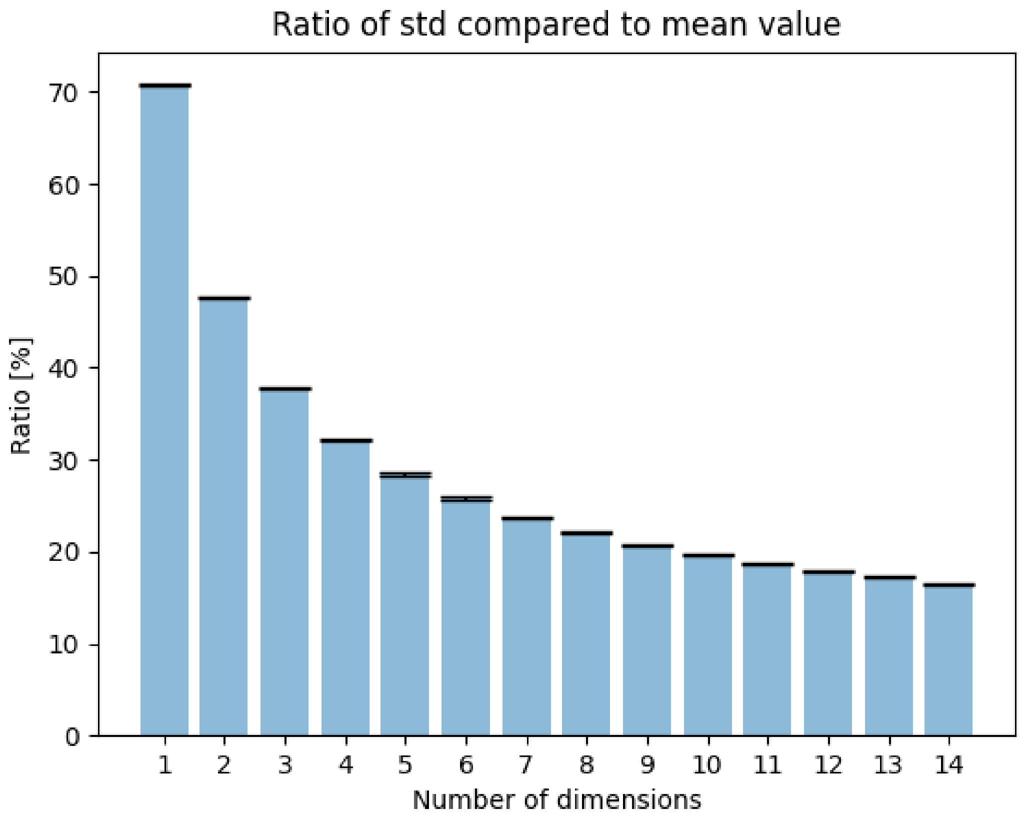
Wykresy przedstawiające średnią odległość pomiędzy punktami, odchylenie standardowe oraz stosunek tego odchylenia do średniej dla hipersześcianu o boku 2, w zależności od liczby wymiarów *n*.



Avarage distances between points
distance
3.0
2.5
2.0
1.5
1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Number of dimensions



Standard deviations of distances between points
standard deviations
0.505
0.500
0.495
0.490
0.485
0.480
0.475
0.470
0.465
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Number of dimensions

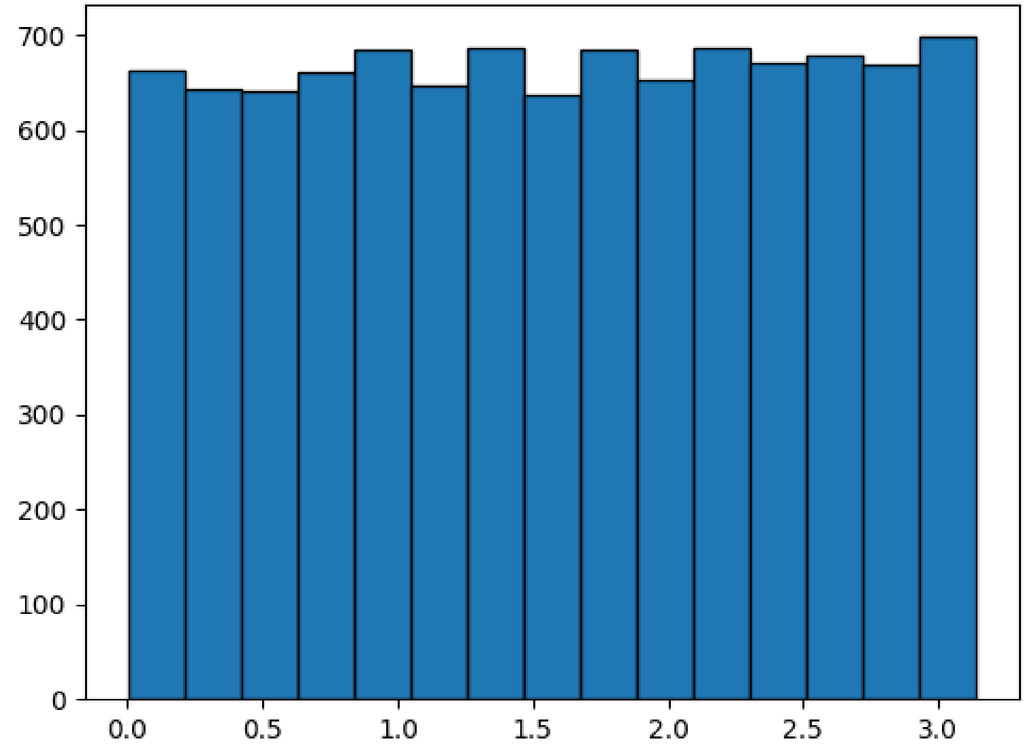


Komentarz:
Średnia odległość między punktami rośnie mniej więcej liniowo w zależności od liczby wymiarów, co więcej, odchylenie standardowe z 10 pomiarów jest bardzo małe, co sugerowałoby, że ta wartość jest w przybliżeniu stała dla danego wymiaru i boku hipersześcianu. Natomiast odchylenie standardowe tych odległości dla pierwszego wymiaru jest wyraźnie niższe niż dla drugiego wymiaru, ale dla wyższych wymiarów nie różni się tak bardzo. Efektem powyższej obserwacji jest wykres nr 3 - stosunek odchylenia standardowego do średniej odległości między punktami wyraźnie maleje wraz ze wzrostem liczby wymiarów. Można z tego wywnioskować, że odległości między punktami dla wyższych wymiarów są dużo bliższe stałej wartości, czyli punkty są bardziej równomiernie rozłożone w przestrzeni.

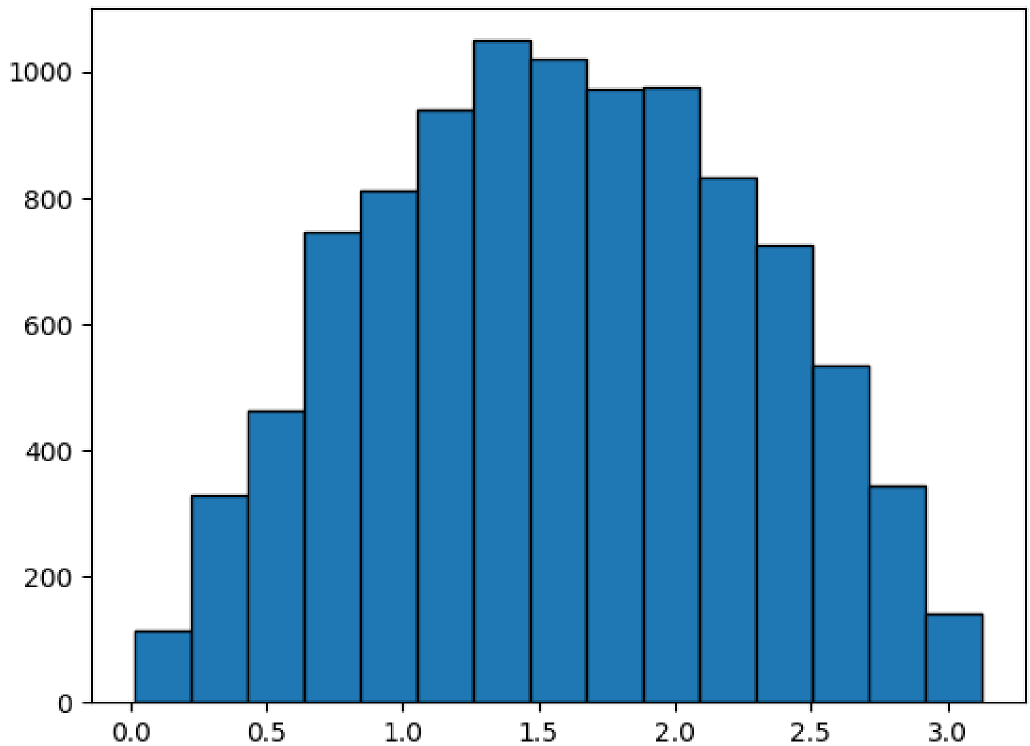
Zad 3.

Dla tego zadania wygeneruję 100000 punktów z hipersześcianu a następnie 10000 razy wylosuję dwie pary punktów i obliczymy kąt, pomiędzy dwoma wektorami. Histogramy kątów pokażę dla wymiarów 2, 3, 5, 10.

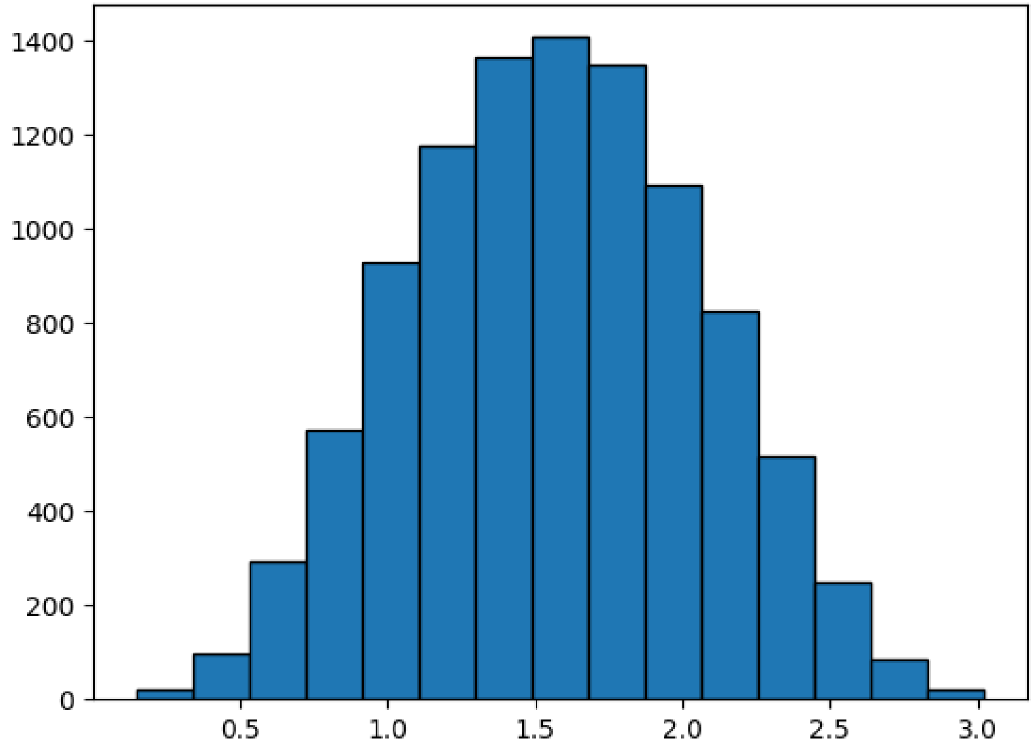
Histogram of angles between 2 random vectors for 2 dimensions



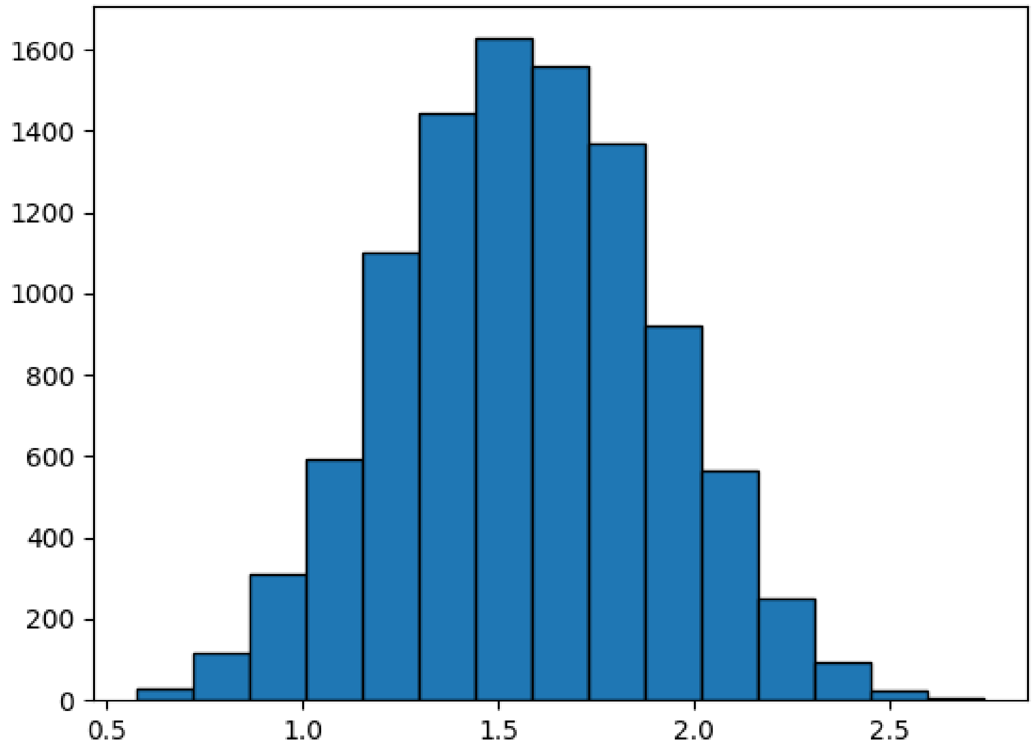
Histogram of angles between 2 random vectors for 3 dimensions



Histogram of angles between 2 random vectors for 5 dimensions



Histogram of angles between 2 random vectors for 10 dimensions



Komentarz:
Dla 2d rozkład wartości kątów w przedziale (0, π) przypominał jednostajny. Jednak wraz ze wzrostem wymiarów, rozkład coraz bardziej zbliżał się do rozkładu normalnego z wartością oczekiwaną równą $\pi/2$. Oznacza to, że im większy mamy wymiar, tym również większe prawdopodobieństwo, że dwa losowe wektory są do siebie prostopadłe.

Wnioski

Liczba wymiarów większa niż 3 jest często trudna do wyobrażenia i jeszcze trudniejsza do zwizualizowania. Jednak istnieją pewne zależności dotyczące punktów w przestrzeni dla większych wymiarów, których warto być świadomym przy operacji na danych, projektowaniu algorytmów i modeli dla danych cechujących się wielowymiarowością.