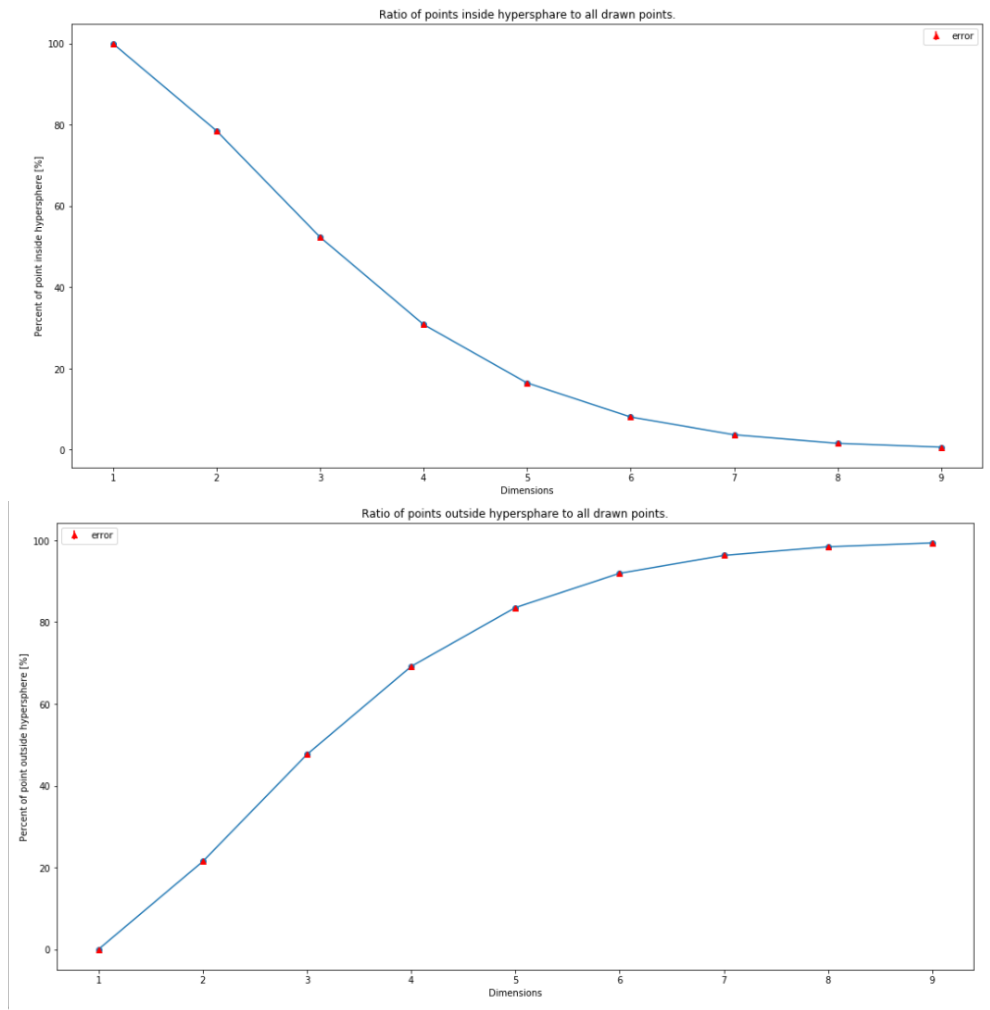


**Raport – Klątwa wymiarów**  
**Podstawy nauczania maszynowego**  
**Wyk. Mateusz Woś**

**Zadanie a.**



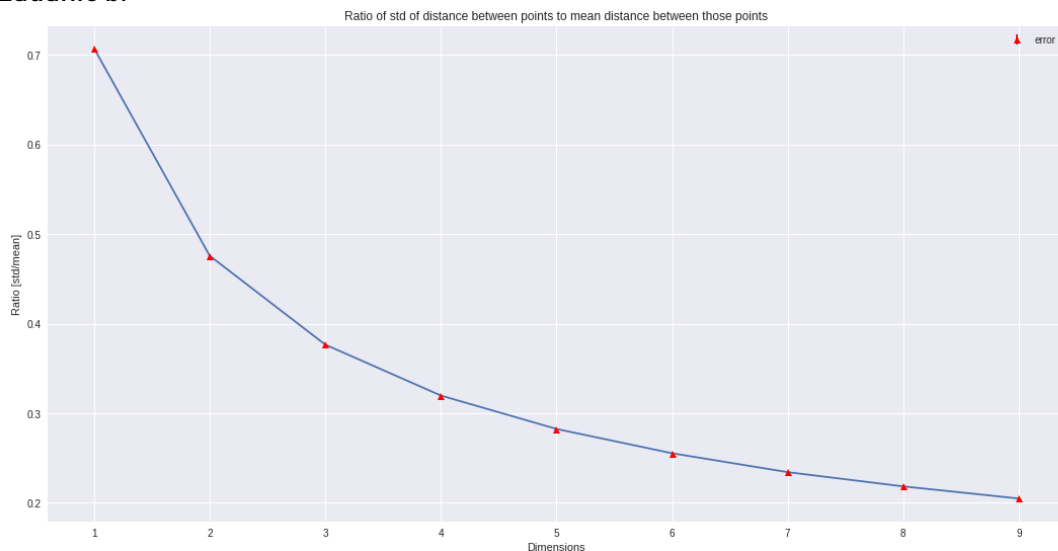
Dla każdego  $n$ -wymiarowego hipersześcianu o boku 1 w którego została wpisana  $n$ -wymiarowa hiperkula o promieniu 1 wygenerowałem 2500000 punktów. Za pomocą metody monte carlo wyznaczyłem stosunek ilości punktów w środku i na zewnątrz hiperkuli do wszystkich wylosowanych punktów.

Powyższe wykresy prezentują stosunek punktów w i na zewnątrz hiperkuli do wymiaru hiperkuli.

Błąd (odchylenie standardowe) był bardzo mały.

Z wykresów można zauważyć, iż wraz ze wzrostem wymiaru spada liczba punktów znajdujących się w środku hiperkuli, punkty znajdują się coraz dalej od kuli. Wysoko wymiarowa hiperkula staje się zbiorem ostrych rogów. Duża części objętości przez to jest poza hipersześcianem.

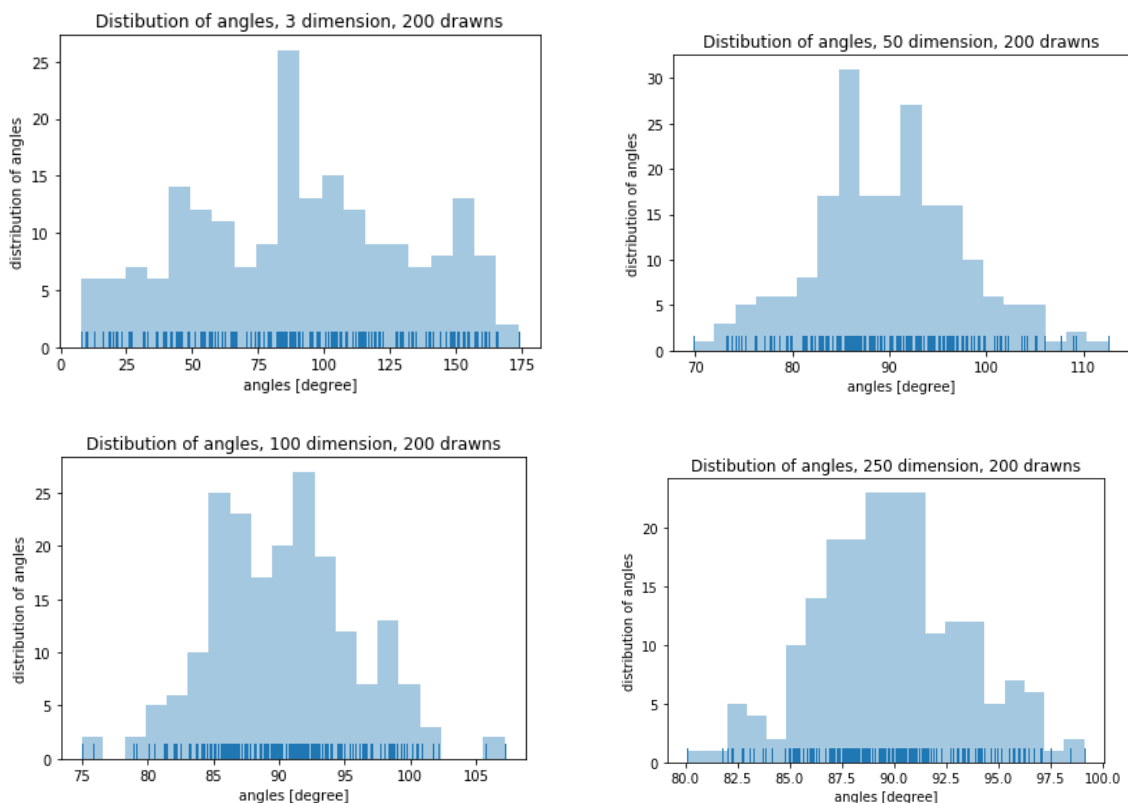
## Zadanie b.



Do wyliczenia danych posłużyłem się google colaboratory, na osobistym komputerze brakowało mi pamięci RAM do ich wyliczenia.

Dla każdego wymiaru hipersześcianu wylosowałem 10k punktów. Na podstawie tych danych wyliczyłem stosunek standardowego odległości między tymi punktami do średniej odległości między nimi. Stosunek ten malał wraz ze wzrostem wymiaru.

## Zadanie c.



Dla każdego z wymiaru(3,50,100,250) wylosowałem po 800 punktów z wygenerowanych wcześniej danych, bez zwracania. Z wylosowanych punktów utworzyłem 400 wektorów, dzięki czemu otrzymałem data set składający się z 200 kątów. Na wykresach można zaobserwować, iż ze wzrostem wymiaru rozkład kątów zbliża się dość znacząco do wartości 90 stopni.

**Wnioski:**

- Wzrost wymiaru -> przestrzeń staje się pustsza. Widać to dobrze na przykładzie wielowymiarowej kuli.
- Wzrost wymiaru -> duży procent danych znajduje się w rogach przestrzeni -> trudniej sklasyfikować dane.
- Wraz ze wzrostem ilości cech danych trudno nam stwierdzić które są z nich bezużyteczne. Niepotrzebne wymiary mogą sprawić, że nasz model będzie działał o wiele gorzej, sam model nie będzie wiedział, które cechy są dla niego niepotrzebne
- Zbyt dużo wymiarowe dane -> będziemy potrzebować o wiele większy data set aby zauważyć wzorzec którego szukamy.