Akademia Górniczo-Hutnicza
Wydział Informatyki, Elektroniki i Telekomunikacji

Łukasz Strzałka

Laboratorium 1 - Klątwa Wymiaru

Zadanie A

Dane

Liczba Wymiarów: **20**
Liczba Punktów: **100000**
Długość Krawędzi Sześcianu: **1000**
Liczba Powtórzeń: **10**

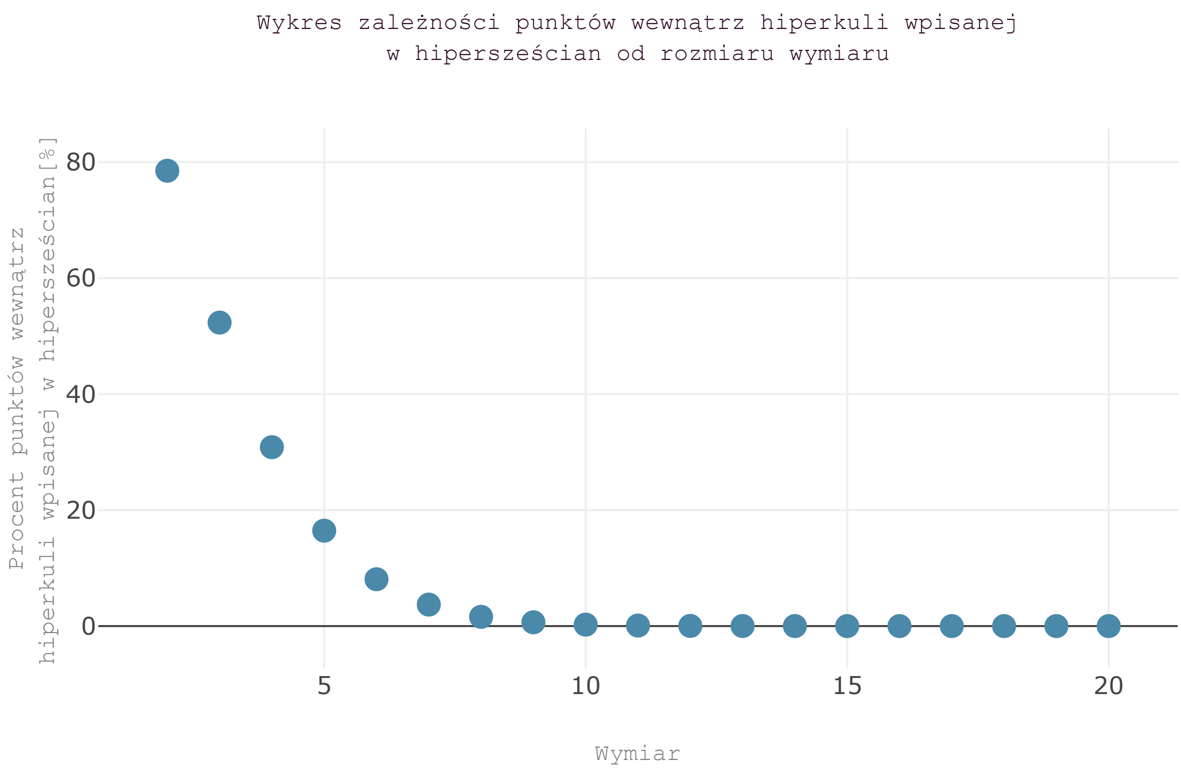
Wykonanie

Program generował zadaną liczbę punktów, żeby następnie zaraz po wygenerowaniu wyliczyć odległość euklidesową między nim a środkiem hiperkuli. Jeżeli ta odległość była mniejsza bądź równa promieniowi hiperkuli to zostawał zwiększany licznik punktów wewnętrznych. Następnie czyyność ta została powtórzona o zadaną liczbę i na poniższym wykresie został przedstawiony średni wynik tych operacji.

Wnioski

Na poniższym wykresie możemy zaobserbować że stosunek ilości punktów znajdujących się w środku hiperkuli do wszystkich punków w hiperszcześcianie maleje początkowo gwałtownie względem wzrostu wymiaru, od pewnego wymiaru już nie obserwujemy tak znacznego spadku. Przypomina to funkcję wykładniczą. Możemy również wnioskować że ten stosunek będzie zmierzał do 0. Error bary są na tyle małe że są niewidoczne, można przybliżyć wykres w wersji HTML.

Wykres



Zadanie B

Dane

Liczba Wymiarów: **20**
Liczba Punktów: **100**
Długość Krawędzi Sześcianu: **1000**
Liczba Powtórzeń: **10**

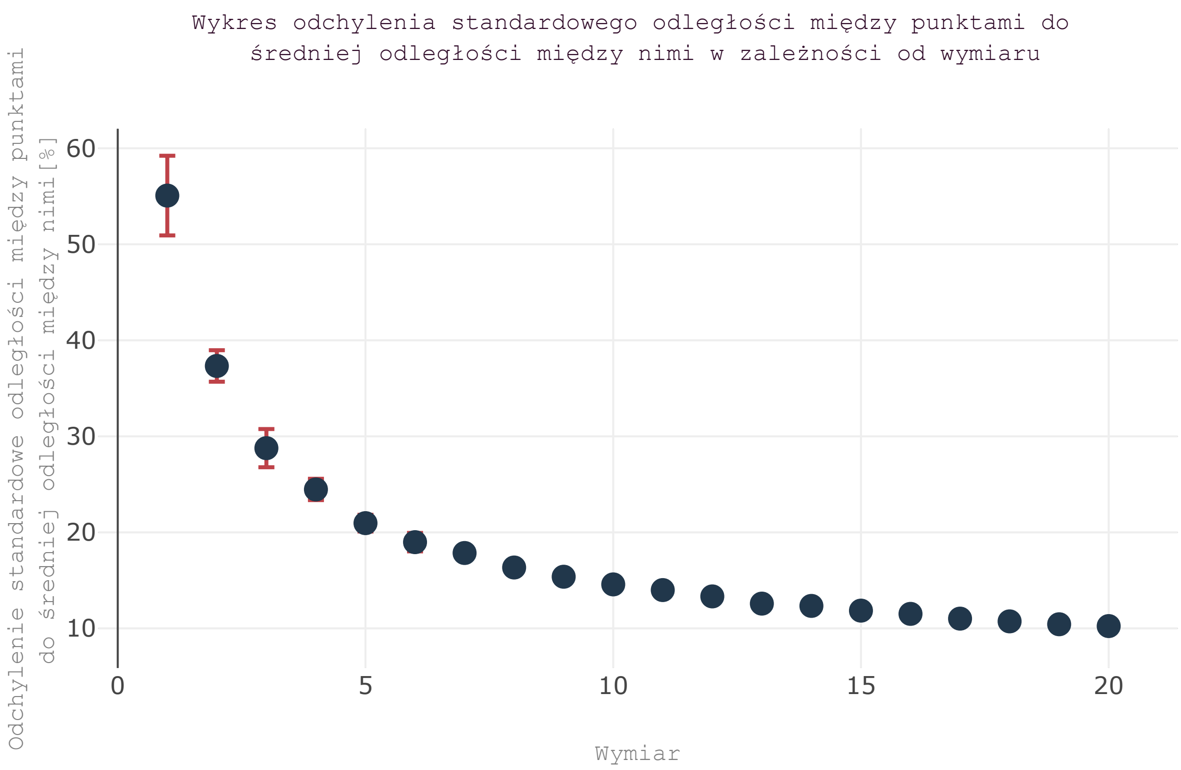
Wykonanie

Program generował zadaną liczbę punktów, następnie zapisywał te punkty do zmiennej. Po zapisaniu wszystkie zadancyh punktów obliczał odległości między wszystkimi punktami i zapisywał je do zmiennej i usuwał same punkty. Następnie liczył standatdowe odchylenie tych odległości i dzielił je przez średnią odległość. Operację tą powtarzał dla każdego wymiaru zadaną liczbę razy.

Wnioski

Możemy zaobsebować na poniższym wykresie że stosunek odchylenia standrdowego odległości między punktami do średniej odległości między nimi maleje względej wzrostu wymiaru gwałtownie do pewnego punktu, następnie od niego maleje tylko nieznacznie. Wykres przypomina funkcję wykładcznią. Obserwujemy również że stosunek ten zmierza do 0. Error bary są na tyle małe że są niewidoczne, można przybliżyć wykres w wersji HTML.

Wykres



Wnioski

Wnioski

Powyższe dwa wykresy są w oczywiste sposób analogiczne. Oba przypominają funkcje wykładnicze i oba zmierzają do 0. Możemy się zastanowić dlaczego tak się dzieje. W pierwszym przypadku liczymy odległość między punktem a środkiem hiperkuli. Objętość hipersześcianu liczymy ze wzoru a^{dim}, a objętość hiperkuli liczymy wzorem:

$$V_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n = \begin{cases} \frac{\pi^k}{k!} \cdot r^n & \text{dla } n = 2k, \\ \frac{2^k \pi^{k-1}}{n!!} \cdot r^n & \text{dla } n = 2k-1, \end{cases}$$

Zauważamy że dzielenie przez silnię wymiaru znacznie zmniejsza jej objętość wzgledem hipersześcianu którego objętość rośnie wykładniczo. W drugim przypadku mamy podobną sytuację, przez dodawanie kolejnych wymiarów odległości między punktami rosną, a odchylenie tych odległości się nie zmienia. Jakie konsekwencje mają uzyskane wyniki dla algorytmów rozpoznawania wzorców? Weźmy życiowy przykład. Mamy serwis randkowy, który zbiera dane o osobach zarejestrowanych i następnie próbuje znaleźć inne osoby pasujące do tej osoby. Czyli próbuje dopasować wzorce. Mamy wyuczony algorytm dla pewnego zestawu danych i daje nam dobre wyniki, użytkownicy się cieszą bo znajdują swoją drugą połowę, właściciel serwisu się cieszy bo dużo ludzi korzysta z jego serwisu. Wpadł na pomysł że więcej osób będzie korzystać z tego serwisu, jeżeli jeszcze zwiększy liczbę informacji żeby jeszcze lepiej dopasować osoby. Niestety, nie uwzględnił w tych planach większej ilości danych potrzebynch do nauczenia systemu i nauczył go na tej samej licznie danych co ostatnio powiększonych o te dodatkowe informacje. Efekt? Dopasowania stają się mniej trafne, użytkownicy nie są zadowoleni, nie używają serwisu, firma upada. Wniosek? Musimy zwiększać ilość danych uczących, wnisokując po wykresach musimy to zrobić wykładniczo.

Wykres