2020-1 Algorithm

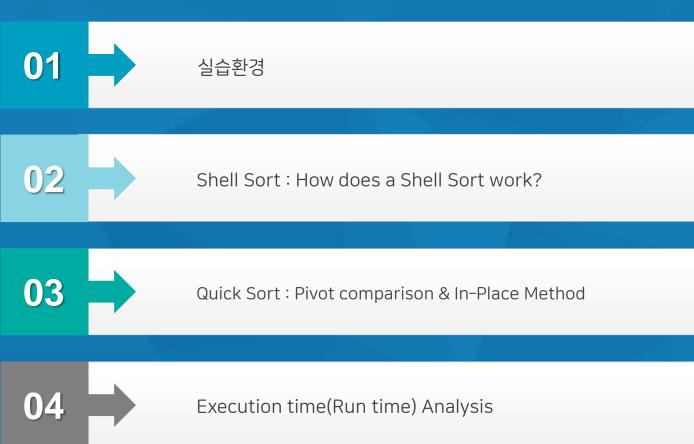
Quick Sort & Shell Sort Comparison

2016270265 이종헌





Contents



01. Development Environment

Win10 Pro, i5-6600, 8GB RAM, Python, Jupyter Notebook



01. Development Environment



Why python and Jupyter Notebook?

- 1. 프로젝트의 진행을 정확한 그래프를 통해 시각적으로 보여드리고 싶었기 때문입니다. matplot을 사용하여 빈도수를 체크하는 코드의 경우 pycharm 혹은 spyder와 같은 인터프리터에서는 결과물밖에 보여지지 않습니다. 쥬피터노트북을 사용함으로 프로젝트의 단계별 진행을 보일 수 있었습니다.
- 2. 모두 오름차순으로 정렬.

02. Shell Sort Algorithm

It is a generalized version of insertion **sort**. In **shell sort**, elements at a specific interval are **sorted**.





안전 정렬 (Stable)	불안전 정렬 (not stable)	
같은 값(key)의 위치가 정렬 과정에서 바뀌지 않는 것	같은 값(key)의 위치가 정렬 과정에서 바뀌는 것	
내부 정렬 (Internal sorting)	외부 정렬 (External sorting)	
데이터의 크기가 주 기억장소 용량보다 적을 경우 기억장소를 활용하여 정렬	데이터의 크기가 주 기억장소 용량보다 클 경우 외부 기억장치를 사용하여 정렬	



What is the Shell Sort?

- 1. Donald L.Shell이 제안한 방법. 삽입정렬을 보완한 알고리즘.
- 2. 삽입정렬이 어느 정도 정렬이 된 상태에서는 대단히 빠르다는(Best일 때 n) 것에서 착안함.
 - 1. 삽입정렬의 문제점 : 요소들이 삽입될 때, 이웃한 위치로만 이동. 멀리 떨어진곳에 삽입되어야한다면 많은 이동을 해야한다.

3. Insertion Sort와는 다르게 Shell Sort는 전체의 리스트를 한번에 정렬하지 않는다.

"It is a generalized version of insertion sort.

In Shell Sort, elements at a specific interval are sorted."



How does a Shell Sort work?

- 1. 먼저 정렬해야 할 리스트를 일정한 분류 기준에 따라 분류 (gap으로 지정)
- 2. 연속적이지 않은 여러 개의 부분 리스트를 생성
- 3. 각 부분 리스트를 삽입 정렬을 이용하여 정렬
- 4. 모든 부분 리스트가 정렬되면 다시 전체 리스트를 더 좁은 간격을 사용하여 적은 개수의 부분 리스트로 만들어 반복
- 5. 간격이 1이 될 때 까지 반복하여 마지막에는 insertion Sort로 정렬

"It is a generalized version of insertion sort.

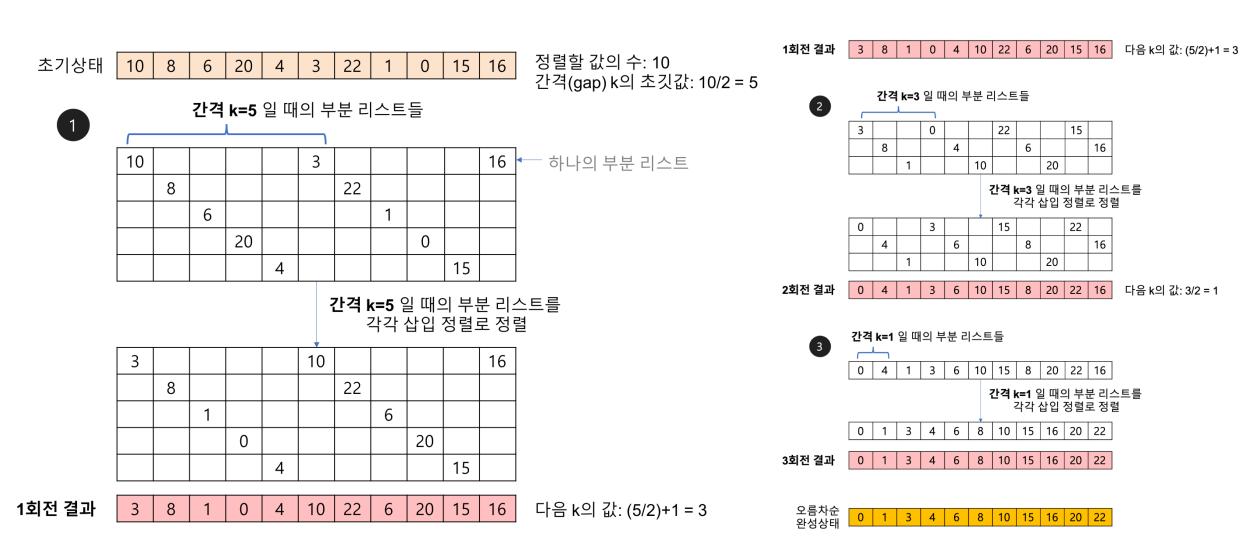
In Shell Sort, elements at a specific interval are sorted."



How does a Shell Sort work?

- 1. 정렬해야 할 리스트의 각 gap번째 요소를 추출하여 부분리스트를 만든다. 요소 사이 간격gap 이라고 한다.
 - l. 간격의 초기값은 정렬할 요소들의 개수 /2
 - II. 생성된 부분 리스트의 개수는 gap의 값과 같다.
- 2. 반복 할 때 마다 간격을 절반으로 줄이면서 부분 리스트에 속한 값들의 개수는 증가한다.
 - I. 간격은 홀수로 하는 것이 좋다.
 - Ⅱ. 간격을 절반으로 줄일 떄 짝수가 나오면 +1을 해서 홀수로 만든다.
- 3. 간격 gap이 1이 될 때까지 반복한다.







파이썬으로 구현한 Shell Sort 함수

```
def gapInsertionSort(arr, start, gap):
       for target in range(start+gap, len(arr), gap):
           val = arr[target]
           i = target
           while i > start:
               if arr[i-gap] > val:
                   arr[i] = arr[i-gap]
               else:
                   break
               i -= gap
           arr[i] =_val
        shellSort(arr):
       gap = Ien(arr) // 2
       while gap > 0:
16
           for start in range(gap):
               gapInsertionSort(arr, start, gap) 🗸
```

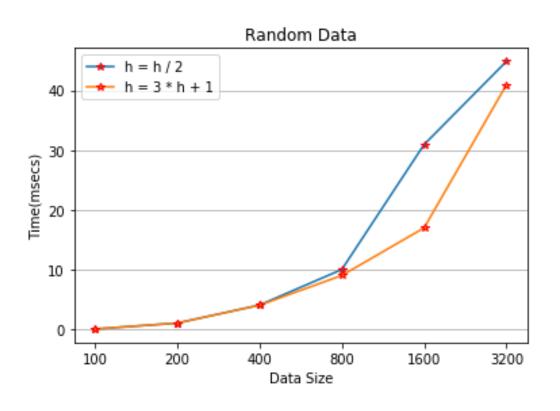
```
def gapInsertionSort(arr, start, gap):
       for target in range(start+gap, len(arr), gap):
           val = arr[target]
            i = target
            while i > start:
                if arr[i-gap] > val:
                   arr[i] = arr[i-gap]
               else:
                   break
                i -= gap
            arr[i] = val
        shellSort2(arr):
       gap = 1
       while gap < len(arr):
               gap = 3*gap +1
       while gap > 0:
19
            for start in range(gap):
               gapInsertionSort(arr, start, gap) 🗸
           gap = gap // 3
```

Why gap = 3*gap + 1?



OEIS	General term $(k \ge 1)$	Concrete gaps	Worst-case time complexity	Author and year of publication
	$\left\lfloor rac{N}{2^k} ight floor$	$\left\lfloor \frac{N}{2} \right\rfloor, \left\lfloor \frac{N}{4} \right\rfloor, \dots, 1$	$\Theta\left(N^2 ight)$ [e.g. when N = 2^p]	Shell, 1959 ^[4]
	$2\left\lfloor rac{N}{2^{k+1}} ight floor+1$	$2\left\lfloor rac{N}{4} ight floor+1,\ldots,3,1$	$\Theta\left(N^{\frac{3}{2}}\right)$	Frank & Lazarus, 1960 ^[8]
A000225	2^k-1	1, 3, 7, 15, 31, 63,	$\Theta\left(N^{\frac{3}{2}}\right)$	Hibbard, 1963 ^[9]
A083318	$2^k + 1$, prefixed with 1	1, 3, 5, 9, 17, 33, 65,	$\Theta\left(N^{\frac{3}{2}}\right)$	Papernov & Stasevich, 1965 ^[10]
A003586	Successive numbers of the form $2^p 3^q$ (3-smooth numbers)	1, 2, 3, 4, 6, 8, 9, 12,	$\Theta\left(N\log^2N\right)$	Pratt, 1971 ^[1]
A003462	$rac{3^k-1}{2}$, not greater than $\left\lceil rac{N}{3} ight ceil$	1, 4, 13, 40, 121,	$\Theta\left(N^{\frac{3}{2}}\right)$	Knuth, 1973, ^[3] based on Pratt, 1971 ^[1]
A036569	$egin{aligned} &\prod_I a_q ext{, where}\ &a_q=\min\left\{n\in\mathbb{N}:n\geq \left(rac{5}{2} ight)^{q+1}, orall p{:}\ 0\leq p< q\Rightarrow \gcd(a_p,n)=1 ight\}\ &I=\left\{0\leq q< r\mid q eqrac{1}{2}\left(r^2+r ight)-k ight\}\ &r=\left\lfloor\sqrt{2k+\sqrt{2k}} ight floor \end{aligned}$	1, 3, 7, 21, 48, 112,	$O\left(N^{1+\sqrt{rac{8\ln(8/2)}{\ln(N)}}} ight)$	Incerpi & Sedgewick, 1985, ^[11] Knuth ^[3]
A036562	$4^k + 3 \cdot 2^{k-1} + 1$, prefixed with 1	1, 8, 23, 77, 281,	$O\left(N^{\frac{4}{3}}\right)$	Sedgewick, 1982 ^[6]
A033622	$\left\{ egin{array}{ll} 9\left(2^k-2^{rac{k}{2}} ight)+1 & k ext{ even,} \ 8\cdot 2^k-6\cdot 2^{(k+1)/2}+1 & k ext{ odd} \end{array} ight.$	1, 5, 19, 41, 109,	$O\left(N^{\frac{4}{3}}\right)$	Sedgewick, 1986 ^[12]
	$h_k = \max\left\{\left\lfloor rac{5h_{k-1}}{11} ight floor, 1_0 = N ight.$	$\left\lfloor \frac{5N}{11} \right\rfloor, \left\lfloor \frac{5}{11} \left\lfloor \frac{5N}{11} \right\rfloor \right\rfloor, \dots, 1$	Unknown	Gonnet & Baeza-Yates, 1991 ^[13]
A108870	$\left\lceil \frac{1}{5} \left(9 \cdot \left(\frac{9}{4} \right)^{k-1} - 4 \right) \right\rceil$	1, 4, 9, 20, 46, 103,	Unknown	Tokuda, 1992 ^[14]
A102549	Unknown (experimentally derived)	1, 4, 10, 23, 57, 132, 301, 701	Unknown	Ciura, 2001 ^[15]







Shell Sort Algorithm의 장점

- 연속적이지 않은 부분 리스트에서 자료의 교환이 일어나면 더 큰 거리를 이동한다. 따라서 교환되는 요소들이 삽입 정렬보다는 최종 위치에 있을 가능성이 높아진다. 배열 뒷부분의 작은 숫자를 앞부분으로 '빠르게' 이동시키고, 동시에 앞부분의 큰 숫자는 뒷부분으로 이동시킨다.
- 부분 리스트는 어느 정도 정렬이 된 상태이기 때문에 부분 리스트의 개수가 1이 되게 되면 셸 정렬은 기본적으로 삽입 정렬을 수행하는 것이지만 삽입 정렬보다 더욱 빠르게 수행된다.
- 알고리즘이 간단하여 프로그램으로 쉽게 구현할 수 있다.



Shell Sort Algorithm의 장점

- 입력 크기가 매우 크지 않은 경우에 매우 좋은 성 능을 보인다.
- 임베디드(Embedded) 시스템에서 주로 사용되는 데, Shell Sort의 특징인 간격에 따른 그룹 별 정렬 방식이 H/W 로 정렬 알고리즘을 구현하는데 매우 적합하기 때문이다.

uClibc: 임베디드 리눅스 전용으로 만들어진 소형 C표준 라이브러리.

"Shell Sort is used in the uClibc library. In the past, ShellSort was used in the Linux Kernel"

03. Quick Sort Algorithm

Quicksort is a divide-and-conquer algorithm.

It works by selecting a 'pivot' element from the array and partitioning the other elements into two sub-arrays, according to whether they are less than or greater than the pivot.





What is the Quick Sort?

1. Divide and Conquer 방식으로 큰 데이터를 작게 쪼개서 하나씩 해결하는 방법을 씀. pivot을 잡아 pivot보다 작은 값은 왼쪽으로, 큰 값은 오른쪽으로 이동을 하게 함.

2. 파이썬의 list.sort() 함수나 자바의 Arrays.sort()처럼 프로그래밍 언어 차원에서 기본적으로 지원되는 내장 정렬 함수는 대부분은 퀵 정렬을 기본으로 합니다.

3. 일반적으로 원소의 개수가 적어질수록 나쁜 중간값이 선택될 확률이 높아지기 때문에, 원소의 개수에 따라 퀵 정렬에 다른 정렬을 혼합해서 쓰는 경우가 많습니다.



How does the Quick Sort work?

- 1. 리스트 요소들 중 하나를 피벗으로 선택하고, 피벗 값보다 작은 값, 동일한 값 그리고 큰 값을 담아둘 3 개의 리스트를 생성합니다.
- 2. 반복문을 통해서 각 값을 피벗과 비교 후에 해당하는 리스트에 추가시킵니다.
- 3. 작은 값과 큰 값을 가지고 있는 배열(리스트)를 대상으로 Quick Sort를 재귀적으로 호출합니다.
- 4. 마지막으로 재귀 호출의 결과를 다시 크기 순으로 합쳐서 정렬된 리스트를 얻을 수 있다.



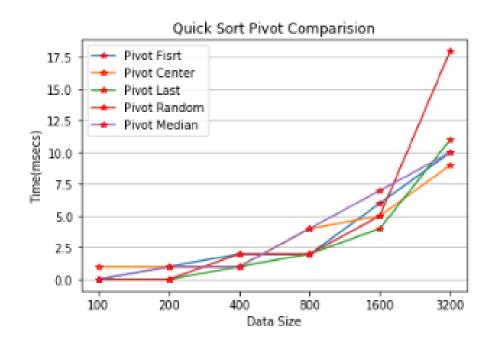
```
def quickSort1(arr): #피빗 처음 값
       if len(arr) <= 1:</pre>
          return arr
      # pivot = arr[len(arr) // 2]
       pivot = arr[0]
       lesser_arr, equal_arr, greater_arr = [], [], []
9
       for num in arr:
10
           if num < pivot:</pre>
11
              Tesser_arr,append(num) #작은 리스트로 보내기
12
           elif num > pivot:
              greater_arr.append(num) #큰 리스트로 보내기
13
14
          else:
15
              equal_arr,append(num)
16
      #작은 리스트와 큰 리스트끼리 다시 제귀하여 뤽정털
17
       return quickSort1(lesser_arr) + equal_arr + quickSort1(greater_arr)
19
   def quickSort2(arr): #피빗 중간 값
21
22
       if len(arr) <= 1:</pre>
          return arr
24
       pivot = arr[len(arr) // 2]
       lesser_arr, equal_arr, greater_arr = [], [], []
26
27
       for num in arr:
29
           if num < pivot:</pre>
30
              Tesser_arr,append(num) #작은 리스트로 보내기
31
           elif num > pivot:
              greater_arr,append(num) #큰 리스트로 보내기
32
33
34
              equal_arr,append(num)
35
      #작은 리스트의 큰 리스트끼리 다시 제귀하여 뤽징털
       return quickSort2(lesser_arr) + equal_arr + quickSort2(greater_arr)
```

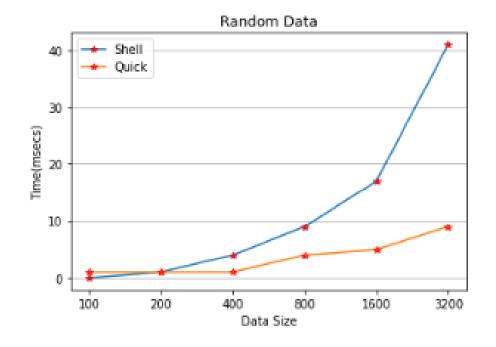
```
if len(arr) <= 1:
    return arr
pivot = arr[random,randint(0,len(arr)-1)] #### 過程 即变 ####
lesser_arr, equal_arr, greater_arr = [], [], []
```

```
def quickSort5(arr): #피빗 처음 중간 비지역의 평균 값

if len(arr) <= 1:
    return arr
pivot = arr[(0+(len(arr)//2)+(len(arr)-1))//3]
```









In-Place Sorting Method

- 정렬 할 원소들을 위한 추가적인 메모리를 사용하지 않고 현재 메모리에서 정렬을 진행
 - 메모리를 아낄 수 있다.
 - 하지만 코드가 조금 더 복잡하다.
 - 왼쪽과 오른쪽 값을 교환하는 과정에서 중복 값의 위치가 바뀔 수 있으므로 unstable하다.



In-Place Sorting Method

```
def quick_sort(arr):
        def sort(low, high):
            if high <= low:
                return
            mid = partition(low, high)
            sort(low, mid - 1)
            sort(mid, high)
 8
10
        def partition(low, high):
11
            pivot = arr[(low + high) // 2]
12
13
            while low <= high:
14
                while arr[low] < pivot:</pre>
15
                     low += 1
16
                while arr[high] > pivot:
17
                    high -= 1
                if low <= high:</pre>
18
                    arr[low], arr[high] = arr[high], arr[low]
19
20
                    low, high = low + 1, high - 1
21
            return low
22
23
        return sort(0, len(arr) - 1)
```

메인 함수인 quick_sort()는 크게 sort()와 partition()로 구성

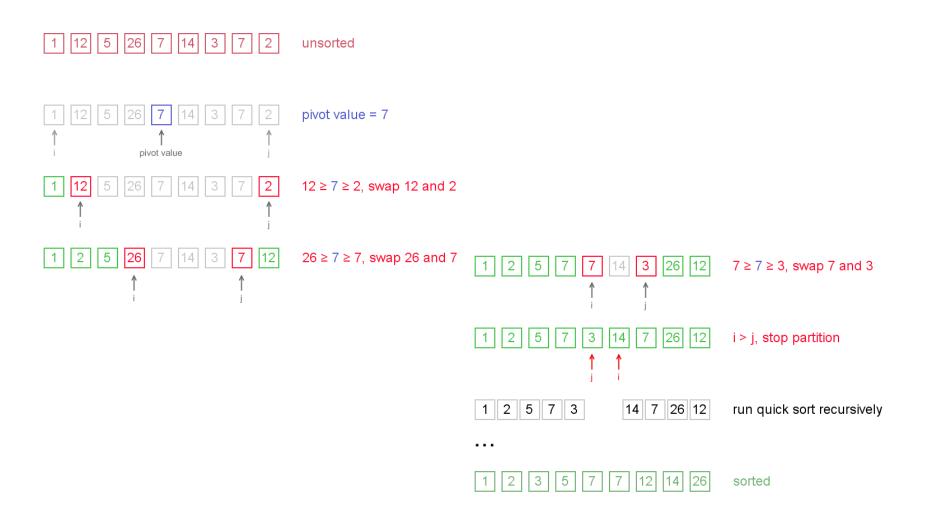
sort() 함수는 재귀 함수이며 정렬 범위를 시작 인덱스와 끝 인덱스로 인자로 받습니다.

partition() 함수는 정렬 범위를 인자로 받으며 좌우측의 값들을 정렬하고 분할 기준점의 인덱스를 return

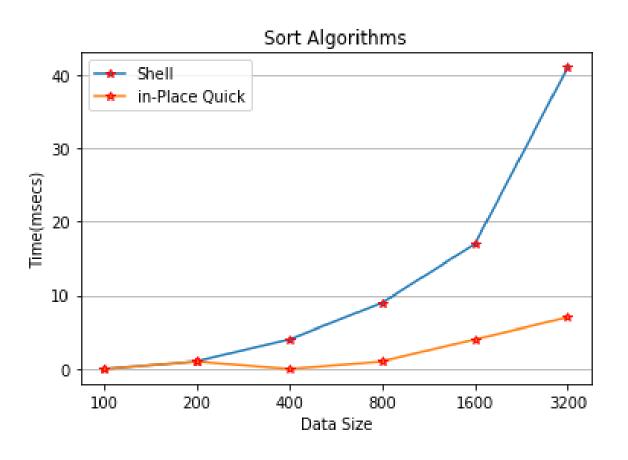
이 분할 기준점(mid)는 sort()를 재귀적으로 호출할 때 우측 리스트의 시작 인덱스로 사용됩니다.

- 리스트의 정 가운데 있는 값을 pivot 값을 선택합니다.
- 시작 인덱스(low)는 계속 증가 시키고, 끝 인덱스(high)는 계속 감소, while 루프로 두 인덱스가 서로 교차해서 지나칠 때까지 반복시킵니다.
 - 시작 인덱스(low)가 가리키는 값과 pivot 값을 비교, 더 작은 경우 반복해서 시작 인덱스 값을 증가시 킵니다. (pivot 값보다 큰데 좌측에 있는 값을 찾기 위해)
 - 끝 인덱스(high)가 가리키는 값과 pivot 값을 비교, 더 작은 경우 반복해서 끝 인덱스 값을 감소시킵니다. (pivot 값보다 작은데 우측에 있는 값을 찾기 위해)
 - 두 인덱스가 아직 서로 교차하지 않았다면 시작 인덱스(low)가 가리키는 값과 끝 인덱스(high)가 가리키는 값을 swap 시킵니다.
- 상호 교대 후, 다음 값을 가리키기 위해 두 인덱스를 각자 진행 방향으로 한 칸씩 이동 시킵니다.
- 두 인덱스가 서로 교차해서 지나치게 되어 while 루프를 빠져나왔다면 다음 재귀 호출의 분할 기준 점이될 시작 인덱스를 리턴합니다.





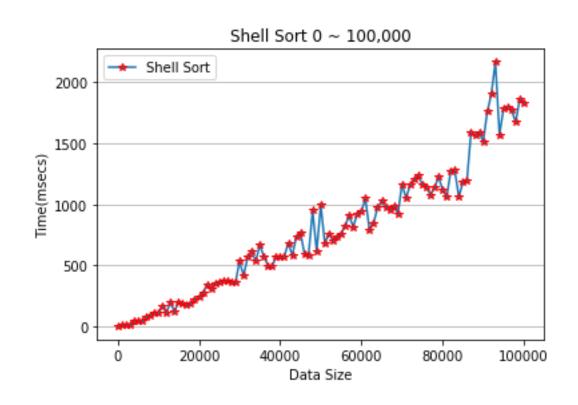


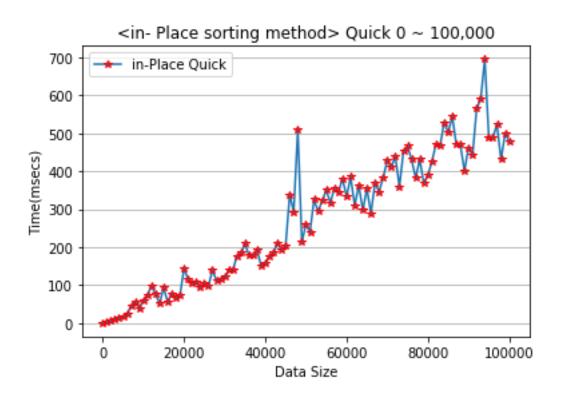


Comparing 2 types of algorithm methods



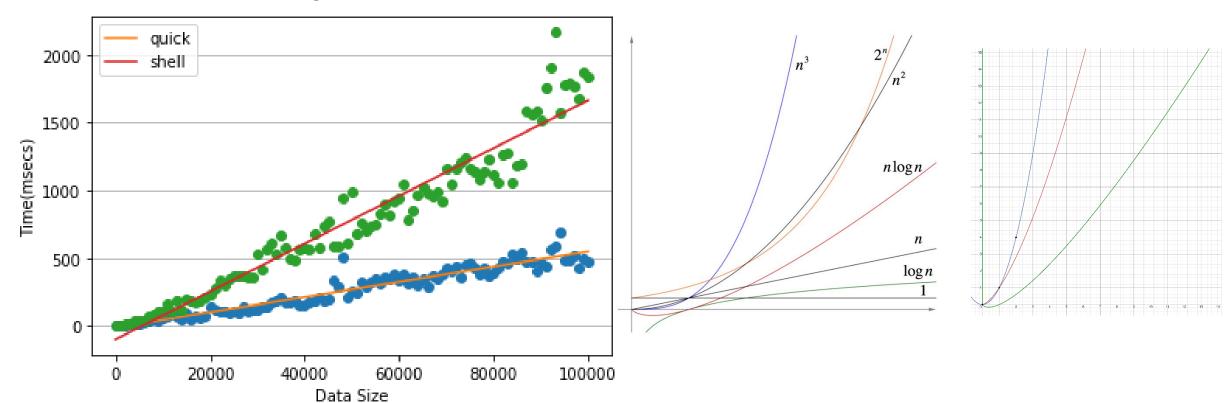








Linear Regression Model





	Worst	Best	Avg
Quick Sort	$O(n^2)$	O(nlogn)	O(nlogn)

	Worst	Best	Avg
Shell Sort	$O(n^2)$	O(n)	$O(n^{1.5})$

THANK YOU

