## 241 assignment data analysis

The number of papers in the pile is denoted as n, whereas the depth denoted as d.

Then depth divide the pile of paper into n/d parts. (assume n/d is integer).

We will analysis the average number of sorting steps needed regarding n and d.

When the depth is d, for the worst scenario, for get a M, we need n/d steps

On averagely:

The first M requires: (1 + 2 + 3 + … + n/d) /(n/d) steps

The second M requires: (1 + 2 + 3 + … + n-1/d) /(n-1/d) steps

…

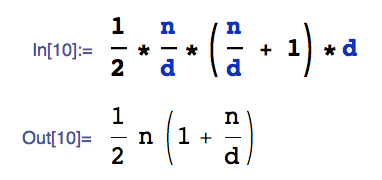
The nth M require 1 steps.

We take laziness, assume all steps need same (1 + 2 + 3 + … + n/d) /(n/d) steps,

then total steps need:

(1 + 2 + 3 + … + n/d) / (n/d) \* n = (1 + 2 + 3 + … + n/d) \* d steps.

Because the arithmetic series (1 + 2 + 3 + … n) = (1/2) \*n\*(n+1), the average number of sorting steps regarding depth d and size n is:

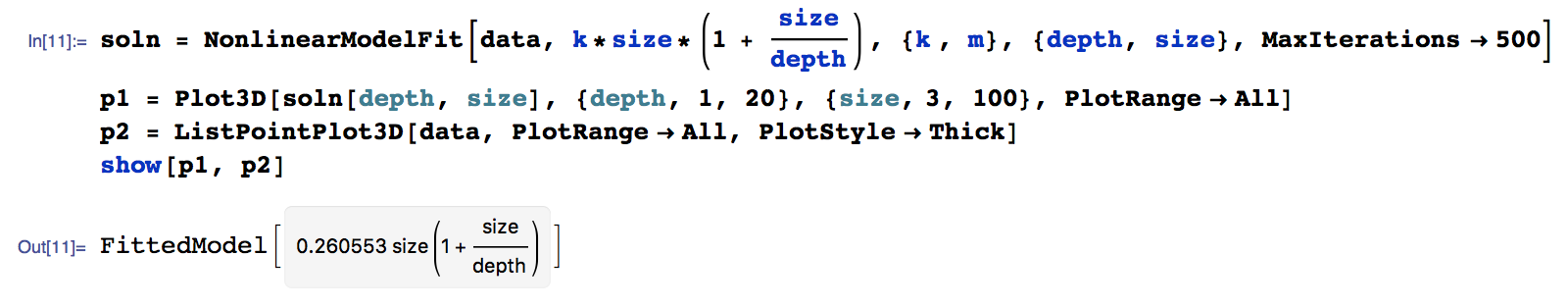


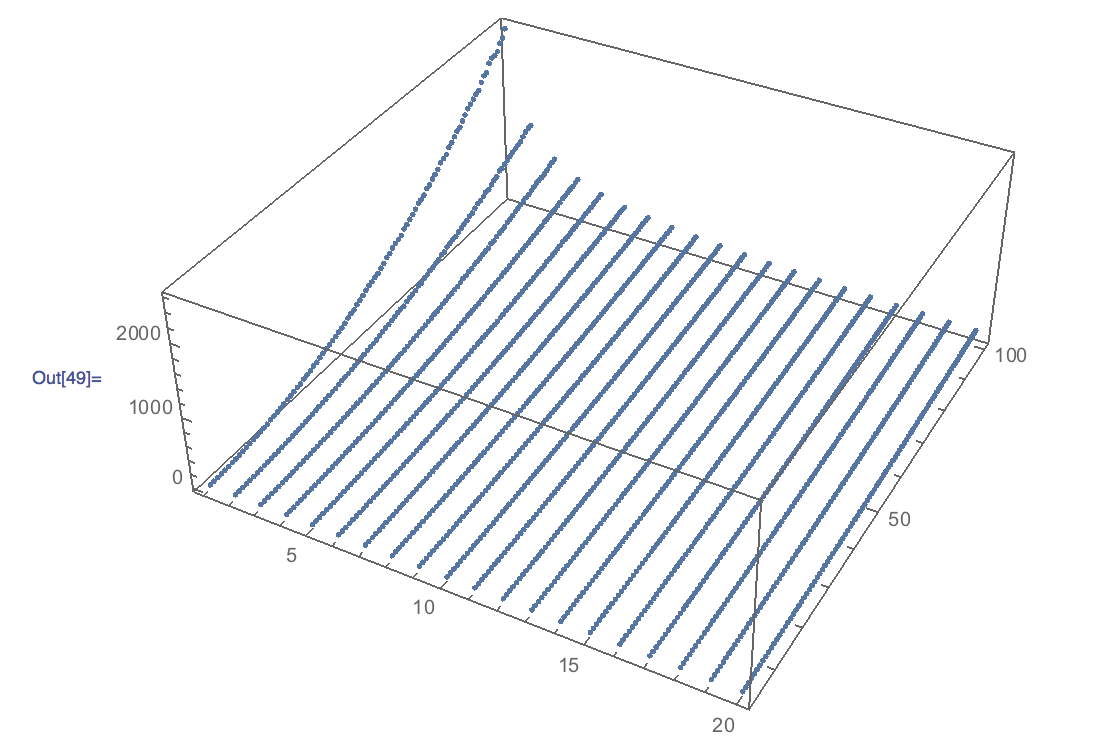
When the depth is 1, the number of sorting steps is O (n^2).

When the depth is n, the number of sorting steps is just n, and output is “MMM…M” with length of n.

When the depth between 1 and n, the sorting steps is between O (n) and O (n^2), is O (n\*(1+n/d)).

We can use regression to find the roughly constant C:



The graph shows the data set plotted with:

d range from 1 to 20

n range from 3 to 100