**Activity 2 a**

As we “scaled up” the word co-occurrence by increasing the number of documents processed from 2 to 54 by using the pairs method, we recorded the performance of the MR infrastructure and plot it as discussed in Chapter 3 of Lin and Dyer’s text.

Upon analysis it can be observed that as the number of documents are increased, we see a low exponential rise in the time taken. The dotted blue line shows the exponential fitting curve.

**Activity 2 b**

We observe a similar behavior on observing recorded the performance of the MR infrastructure 3-gram co-occurrence. However, the time taken for executing 3gram co-occurrence is higher than 2 gram co occurrence

**Activity 2 Analysis**

Average time difference between 2gram and 3-gram co-occurrence for 54 documents is 16ms.

We can see the time difference as the measure of performance for 2gram MR job and 3 gram MR job.

It is evident that for less number of documents the performance is similar, while for large no of documents 2gram co-occurrence takes lot less time than 3 grams because of less time consumption in forming neighbors.