Exercises: segmented least squares

Questions

1. In the lecture I claimed there were $2^{\{N-1\}}$ ways to partition the sequence \{1, 2, ..., N\} into k disjoint sequences (or segments) where, by definition, the segments contain consecutive numbers (or a single number). I refered to this partition as a segmentation. I also presented a recurrence

$$f(N, k) = f(N-1, k-1) + f(N-1, k)$$

for the number of segmentations that contain exactly k segments.

For the number of all possible segmentations, we need to sum $f(N,k)$ over k.

Using the above recurrence and the fact that $f(1,1) = 1$, prove using induction that the number of all possible segmentations is

$$\sum_{k=1}^{N} f(N, k) = 2^{N-1}.$$ 

2. One question that came up in class is whether the recurrence above, and the exponential growth of the number of segmentations, is really a problem in practice since many of the possible segmentations consist of a lot of singletons i.e. segments with just one element.

Maybe the “exponential” growth is due to these cases? I responded that we don’t want to categorically ignore segmentations that contain a segment with a single element, since there might be a perfect fit of a line to the left of some $x_i$ and a perfect fit to the right of $x_i$, and so the $(x_i, y_i)$ might be an outlier and properly segmented as such!

Another way to convince yourself that an exponential growth can arise even in cases in which there are NO segments containing a single element is as follows. Suppose we only consider segmentations that contain $(x_i, x_{i+1})$ pairs, where i is odd and $i+1$ is thus even. That is, $x_1 x_2$ are always in the same segment, $(x_3, x_4)$ are always in the same segment, etc. What can you say about the number of segmentations that satisfies this strong constraint?

Answers

1. To prove the claim using induction, I am going to switch variables and use i instead of k for the number of segments in the segmentation and I’ll use N=K for the induction hypothesis. That way we don’t have more than one k/K floating around to confuse us.
The base case $N=1$ for the proof is obvious, since there is only one way to segment a sequence with one element, i.e.

$$\sum_{i=1}^{1} f(1, i) = f(1,1) = 2^{1-1} = 1.$$

For the induction hypothesis, we let $N=K$ and assume/hypothesize that

$$\sum_{i=1}^{K} f(K, i) = 2^{K-1}.$$  

From the recurrence relation we have

$$\sum_{i=1}^{K} f(K, i) = \sum_{i=1}^{K} f(K-1, i) + \sum_{i=1}^{K} f(K-1, i - 1).$$

We can rewrite the right side as follows. For the first term, there is no way to make $K$ segments out of $K-1$ elements, and so the $i=K$ term is 0. For the second term, we drop the $i=1$ case since there is no way to have a segmentation with 0 segments. Rewriting gives:

$$\sum_{i=1}^{K} f(K, i) = \sum_{i=1}^{K-1} f(K-1, i) + 2 \sum_{i=1}^{K-1} f(K-1, i - 1).$$

Now for the induction step: we write the above result for $N=K+1$ instead of $N=k$,

$$\sum_{i=1}^{K+1} f(K+1, i) = 2 \sum_{i=1}^{K} f(K, i).$$

Applying the induction hypothesis by substituting into the right side gives

$$\sum_{i=1}^{K+1} f(K+1, i) = 2 \times 2^{K-1} = 2^K$$

which is exactly what we need to show.

2. The case described in the question reduces to a problem of half the size i.e. $N/2$ pairs, we ask how many segments we can make from these pairs. The answer follows immediately from the previous result namely $2^{(N/2) - 1}$, or $\sqrt{2}^N - 1$. This is still an exponential growth in the number of segments.