**Foundations of Deep Learning – Homework Assignment #2**Adi Album & Tomer Epshtein

**Part 3: (3)**

Problem:

Consider the following “quadrant” partition of the input elements:

Prove that under this partition, the separation rank of a function realized by a deep network (with hidden layers) is no greater than where stands for the width of layer .

Solution:

First, a visualization of the partition:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

Let be a deep net. Denote by its corresponding tensor given by the HT decomposition:

…

…

Let’s view ’s matricization w.r.t :

{Linearity of matricization and “matricization of outer product equals Kronecker product of matricizations” property}

Let’s take another step back through the HT decomposition:

{Linearity of matricization and “matricization of outer product equals Kronecker product of matricizations” property}

Putting it all together:

Let’s understand that matricizations’ indices:

So

* produces a rank 1 matrix
* A sum of rank 1 matrix produces a ranked matrix

Denote:

So we have , for all .

We have:

We saw in class that for any two matrices & .  
So :

Therefore:

As the sum of matrices with rank .