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

**Part 3: (2)**

Problem:

Modify the expressive efficiency analysis given in class to the case in which the deep network has hidden layers (Assume is a power of ), with pooling windows of size (In class we treated hidden layers with pooling windows of size ).

Solution:

Assume is a power of 4.   
Denote by , the hypothesis space corresponding to a deep network with pooling windows of size and with depth

Similarly to notation in class, we’ll denote its learnable weights by:

Where .

Let’s now state the expressive efficiency statement for this hypothesis class w.r.t , the hypothesis space corresponding to shallow nets:

1. *,*  such that
2. for such that unless .

**Proof 1:**  
Nearly identical to the proof in class for deep nets with windows of size and depth .  
Let . Choose . Let . I.e. is a function realizable by a shallow network of width . We will prove it is realizable by a deep network with window size , depth and width .

Denote by and the learnable weights of the shallow net realizing . ( – Denoting shallow net weights)

We will now define the weights for the deep net so that it realizes :  
With :  
For We define the weights for the conv layers by:

Where is a vector with a single entry “1” at position and in all other entries.  
Under this definition, the 1x1 conv layers act as “passthru” layers, and the pooling layers *together* constitute a global pooling (multiplying 4 entries at a time).

Altering the weights in hidden layer : and output weights we can realize using a deep network.

Define:

Using above weights of layer and output layer , together with “passthru” conv layers out deep-net realizes .

**Proof 2:**

For this part of the proof we will need to use a decomposition of the network’s behavior, similar to the HT decomposition we saw in class for deep nets with windows of size and depth

In class we saw the equivalence of shallow nets to tensors defined by the decomposition, and the equivalence of deep nets (with windows of size and depth ) to tensors defined by the decomposition:

In our case, deep nets with windows of size and depth we obtain the following decomposition which we will call the STAR decomposition denoted by :

…

…

As we saw in class, by viewing the matricization of the CP decomposition (shallow-nets) we saw that for any choice of learnable parameters, any shallow net is equivalent to a tensor such that as a sum of rank 1 matrices.

We’ll consider a similar analysis for STAR decomposition (Denoted by ) corresponding to deep nets with windows of size and depth .

We’ll define the weights of the deep net as follows:

* Layer 0:  
  Let be the width of hidden layer 0. Denote “Layer 0” weights by:  
  And define:  
  for all where is “1” at entry and elsewhere.  
  And:

for all , where is the all zeros vector.

* Layer 1:  
  Let be the width of hidden layer 1. Denote “Layer 1” weights by:

for all where is the all ones vector and is the all zeros vector.  
  
Here we achieve:  
 is the identity matrix for and the zero matrix for .

* Layers 2, …, (L-1):  
  Let be the widths of layers . Denote their weights by:  
  For , , .
* Output weights:  
  Define .

Under our above assignments we achieve with the canonical decomposition:

{Linearity of matricization}

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

… Keep going down the network with (\*) decomposition … Calculations identical to those held in class …

appears times, where is a identity matrix.

In particular we showed a realization of a tensor by decomposition which corresponds to a deep net with windows of size and depth The matricization via the canonical matricization produces a matrix with

**We’ll summarize:** We saw in class that any realization of a shallow net produces a tensor denoted by such that it holds the property: where is shallow net’s width.

We now saw an implementation of a deep net with windows of size and depth such that if we denote it’s corresponding tensor by we have .  
For this tensor to be realized by a shallow net we must have and in particular we must have a network with exponential width (exponential in ).