Homework 02 – Linear Algebra

Arthur J. Redfern
arthur.redfern@utdallas.edu
Jan 23, 2019

---

0 Outline

1 Logistics

Assigned:	Wed Jan 23, 2019
Due:	Wed Jan 30, 2019
Format:	PDF uploaded to eLearning

With respect to the linear algebra slides (https://github.com/arthurredfern/UT-Dallas-CS-6301-CNNs/blob/master/Lectures/xNNs_02_LinearAlgebra.pdf), homework and test questions will focus on the material in the "Matrix transforms" and "Layers built from linear transforms" sections. Material in the previous sections is there such that the slides are a self contained presentation of material.

2 Reading

1. Read: A guide to convolution arithmetic for deep learning
   https://arxiv.org/abs/1603.07285

3 Theory

Matrix multiplication

2. What is the arithmetic intensity for matrix matrix multiplication with sizes M, N and K (BLAS notation)? Prove that arithmetic intensity for matrix matrix multiplication is maximized when M = N = K (all matrices are square).

Dense layers

3. Consider a dense layer that transforms input feature vectors to output feature vectors of the same length ($N_o = N_i$). What is the complexity (MACs and memory) of this layer applied to inputs created from vectorized versions of the following:

MNIST:	1 x 28 x 28
CIFAR:	3 x 32 x 32
ImageNet:	3 x 224 x 224 (typical use)
Quasi 1/4 HD:	3 x 512 x 1024
Quasi HD:	3 x 1024 x 2048

4. In practice, why can't you flatten a quasi HD input image (3 x 1024 x 2048) to a 6291456 x 1 vector and use densely connected layers to transform from data to weak features to strong features to classes?

5. Say I have trained a dense layer for an input of size 1024 x 1. Can this dense layer be applied to an input of size 2048 x 1? What about 512 x 1?

CNN style 2D convolution layers

6. Prove that CNN style 2D convolution with N_o x N_i x F_r x F_c filter, N_i x L_r x L_c input and N_o x (L_r – F_r + 1) x (L_c – F_c + 1) output can be lowered to the sum of (F_r*F_c) matrix matrix multiplications with N_o x N_i matrices made of filter coefficients where the elements of each matrix comes from a single $f_r \in \{0, \ldots, F_r - 1\}$ and single $f_c \in \{0, \ldots, F_c - 1\}$ index and N_i x (M_r*M_c) matrices made from inputs.

7. Consider a CNN style 2D convolution layer with filter size N_o x N_i x F_r x F_c. How many MACs are required to compute each output point?

8. How does CNN style 2D convolution complexity (MACs and memory) scale as a function of
   - Product of the image rows and cols (L_r*L_c)?
   - Product of the filter rows and cols (F_r*F_c), assume N_i and N_o are fixed?
   - Product of the number of input and output feature maps (N_i*N_o)?

9. Consider a CNN style 2D convolution layer with filter size N_o x N_i x F_r x F_c. How many 0s do I need to pad the input feature map with such that the output feature map is the same size as the input feature map (before 0 padding)? What is the size of the border of 0s for $F_r = F_c = 1$? What is the size of the border of 0s for $F_r = F_c = 3$? What is the size of the border of 0s for $F_r = F_c = 5$?

10. Consider a CNN style 2D convolution layer with N_o x N_i x F_r x F_c filter, N_i x L_r x L_c input (L_r and L_c both even) and $P_r = F_r - 1$ and $P_c = F_c - 1$ zero padding. What is the size of the output feature map with striding $S_r = S_c = 1$ (no striding)? What is the size of the output feature map with striding $S_r = S_c = 2$? How does this change the shape of the equivalent lowered matrix equation?

11. Say I have a trained a CNN style 2D convolution layer for an input of size 3 x 1024 x 2048. Can this CNN style 2D convolution layer be applied to an input of size 3 x 512 x 1024? What about 3 x 512 x 512?

RNN layers

12. In a standard RNN, if the state update matrix is constrained to a diagonal, what does this do for the mixing of the previous state with new inputs?

Attention layers

Will defer questions on attention layers for now

Average pooling layers

13. The size of the input to a global average pooling layer is 1024 x 16 x 32. What is the size of the output? What is the complexity (MACs) of the layer?

4 Practice

None