COMP4702/COMP7703/DATA7703 - Machine Learning
Homework 7 - Deep Learning
Solutions

Marcus Gallagher

Core Questions

1. Given the following matrices:

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, K = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

where H is the 5×5 input data (e.g. pixel values) and K is a 2×1 kernel, perform a convolution operation on H using K (stride of 1).

Solution:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

2. Given the following matrices:

$$X = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}, K = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

where X is the 5×5 input data (e.g. pixel values) and K is a 2×2 kernel, perform a convolution operation on X using K (stride of 1).

Solution:

$$\begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 0 & 1 \\ 2 & 0 & 1 & 0 \end{pmatrix}$$

3. Repeat Question 2 but use a stride of 2 and use padding (zero values): add a row to the bottom and a column to the right of X (making it a 6×6 matrix).

   Solution:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

4. Perform 3×3 average pooling on matrix H from Question 1 (stride of 1).

   Solution:

$$\begin{pmatrix} \frac{5}{9} & \frac{3}{9} & \frac{5}{9} \\ \frac{5}{9} & \frac{3}{9} & \frac{5}{9} \\ \frac{5}{9} & \frac{3}{9} & \frac{5}{9} \end{pmatrix}$$

Extension Questions

5. Examine Figures 6.6 and 6.7 in the Deep Learning book (http://www.deeplearningbook.org/contents/mlp.html). In your own words (no more than half a page), explain what the experimental results presented are intended to demonstrate?

   The key point that the Figures are trying to make is that there is something good about deep networks (i.e. many layers). To do this they show that (on some problem) test performance increases as more layers are added. This is contrast to making the network bigger in other ways (Fig.6.7) where we do not see the same effect. Marking notes: Answers no more than half a page. Students should be rewarded for describing this in their own words as opposed to repeating back what is in the book (e.g. figure captions). Reward any relevant reflection or critical analysis of the Figures. For example, some discussion about model complexity. Maybe someone looked at the papers the graphs come from to see what the dataset or experimental details were? Clearly there are also many questions that could be raised about making a general claim from these results (only one dataset? How many trials? What if the curve was extended to more layers? etc.) Any mention of this is good and could contribute to exceeding expectations.