Deep Learning

Problem Sheet 3 —

1. (a) What is a pooling layer in a Convolutional Neural Network? What is the difference between Max pooling and Average pooling?

   (b) What are the stride and padding parameters in a convolutional layer?

   (c) A CNN architecture is described in the table below and takes as input an image and produces a 10-dimensional probability vector and is trained using cross-entropy loss. The architecture consists of max pooling layers as well as convolutional layers.

layer	0	1	2	3	4	5	6	7
type	input	conv	pool	conv	pool	conv	conv	loss
num. filters	-	5x5x1	2x2	5x5x20	2x2	4x4x50	1x1x500	-
stride	-	1	2	1	2	1	-	-
padding	-	0	0	0	0	0	0	-
data shape	1x28x28x1							
receptive field	1							

   The input is a $1 \times 28 \times 28 \times 1$ tensor representing $batch \times width \times height \times channels$. Calculate the data shape and receptive field for each layer.

   (d) Standard precision numbers take up 4 bytes per number. Half precision takes only 2 bytes per number. What are the advantages and disadvantages of using half precision?

   (e) Many CNN training problems are compounded by a lack of available training data. Describe some data augmentation techniques to artificially increase the amount of training data. Some of these techniques are performed at training time on the mini-batch before the forward pass. Suggest why?

2. (a) In a classification problem with three classes, a model will output a 1×3 vector, c, that predicts how strongly the input belongs to a class. However, we want a probability vector, $p \in \mathbb{R}^{3\times1}$. Such a probability vector must sum to 1 and each element must be positive. To ensure this a Softmax function is used which looks like:

$$\textbf{Given } c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \tag{1}$$

$$p_i = \frac{e^{c_i}}{\sum_{k=1}^{3} e^{c_k}} \forall i, k \in 1, 2, 3 \tag{2}$$

Verify whether p fulfils the necessary criteria of all elements being non-negative and summing up to 1.

(b) Compute the derivative $\frac{\partial p_i}{\partial c_j}$. Consider the two scenarios when $i = j$ and $i \neq j$.

(c) Suppose that an image, I is corrupted by white (Gaussian) noise $n \sim N(0, \sigma^2)$ giving us the result $Y = I + n$. One approach to denoising the image is to take N snapshots of an object from the same view, yielding multiple images and then taking the average of all the images. Here, each image is given by $Y_i = I + n_i \forall i \in 1, ..., N$. Taking the average of the N noisy images yields the denoised image $Y_d = \frac{1}{N}\sum_N Y_N$. Derive the mean, μ_d, and variance, σ_d^2, of the denoised image Y_d.

(d) Given an image:

3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65
3	4	8	10	22	45	50	65

And the horizontal and vertical 3×3 Prewitt filter kernels:

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

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

Assuming zero padding, compute the gradient magnitude of the second row after convolving with both kernels.

(e) Repeat the previous question but now use reflection (mirrored) padding.

3. (a) What is the difference between invariance and equivariance? Is convolution either of these?

(b) What is a 2D separable filter? Are all 2D filters separable?

(c) Is the following 2D filter separable?

$$F = \begin{pmatrix} 2 & 3 \\ 1 & 1 \end{pmatrix}$$

If so separate it, otherwise explain why it is not separable.

(d) A CNN has four consecutive 3×3 convolutional layers with stride 1 and no pooling. How large is the support of a neuron in the fourth layer?

(e) Why are skip connections used in deep CNN architectures such as ResNet?

(f) Calculate the (big O) computational complexity of applying both separable and non-separable $K \times K$ Gaussian filters to an $N \times N$ image.