Homework #4: Deep and Reinforcement Learning

Module 2, Lecture 1: Convolutional Networks I

Srikanth Pai, Asst. Professor, MSE

Instructions

Problems 1, 2, 4, and 8 ask for a written explanation; a short paragraph suffices. Problems 3, 5, 6, and 7 are computational; show the arithmetic and state the dimension at each stage. Answers should be concise and precise.

Problem Set

1. Explain why image classification calls for a specialised network design. Why can a plain multilayer perceptron not be used directly on raw pixels? Address both the number of parameters and what an MLP fails to exploit about the structure of an image.

2. Define convolution (equivalently, cross-correlation) for a discrete signal. State the properties of a vision problem that make a convolutional layer a good fit for it. Illustrate each property with a one-dimensional convolution.

3. Define striding and padding, and state what each controls. A grayscale image has size $M \times N$. A convolutional layer applies a filter with kernel size k, padding p, and stride s. Give the dimensions of the output feature map.

4. Define pooling. Explain how a pooling layer contributes to translation invariance, and why this is desirable for classification.

5. Let $x = (1^2, 2^2, 3^2, \ldots, 100^2)$ be the vector of squares of the first 100 natural numbers. Convolve x with the filter $(1, -2, 1)$ using stride 3 and zero-padding 1. Apply max pooling with window size 2 and stride 2 to the result. Compute the final vector.

6. A digital camera produces a 30-megapixel RGB image. The first layer applies 64 different 3×3 filters, with stride 1 and padding chosen so the spatial dimensions are preserved. Ignore bias parameters.

   (a) As a convolutional layer, with the same filter weights shared across all spatial locations, how many trainable weights does the layer have?

   (b) Suppose instead the first layer is fully connected: the image is flattened to a single input vector, and the layer maps it to an output of the same total size as the convolutional layer's output (the full-resolution map of 64 channels). Approximately how many trainable weights would this layer have?

   (c) Compute the ratio of the weight count in the fully connected layer to that in the convolutional layer.

   (d) The convolutional layer differs from the fully connected one in two ways: each output sees only a local patch rather than the whole image, and the patch weights are reused at every location. Explain how each of these two properties contributes to the difference in the weight count.

7. A 30-megapixel RGB image of size $6000 \times 5000 \times 3$ is passed through a 2×2 max-pooling layer with stride 2. Compute the size of the output, and state the main benefits of pooling.

8. Consider a one-dimensional convolution followed by a ReLU nonlinearity. Show that this operation can be written as one layer of a feedforward network. Write out the matrix of the affine transform that the convolution induces. Using this matrix, explain what weight sharing means.