Deep Learning Winter 2025–2026

Homework 2 — Convolutional Networks

Naaman Kopty naamankoptyta@gmail.com

Due: 24.12.2025

Questions 1–2: (70 Points)

Programming assignment — See attached notebook in Moodle.

Question 3: Translational Properties of CNNs (18 Points)

Convolutional neural networks are often motivated by geometric properties of images, in particular the fact that semantic content typically does not depend on exact pixel alignment.

Throughout this question, refer to Figure 1, which displays three spatially shifted versions of the same digit image.

(a) Provide concise and formal definitions of *translational equivariance* and *translational invariance*, expressed using short function notation. Your answer should clearly distinguish between these two properties.

(b) Identify which components of a convolutional neural network equipped with max-pooling exhibit equivariance, and which components introduce invariance. Support your answer by referring to the computational roles of convolution and pooling.

(c) Consider the three shifted images in Figure 1. Discuss whether a CNN with standard max-pooling (e.g., 2×2 or 3×3 pooling with stride > 1) will necessarily produce identical internal activations or class scores for all three versions. Relate your explanation to receptive field growth, discretization effects introduced by pooling stride, and the concepts defined in part (a).

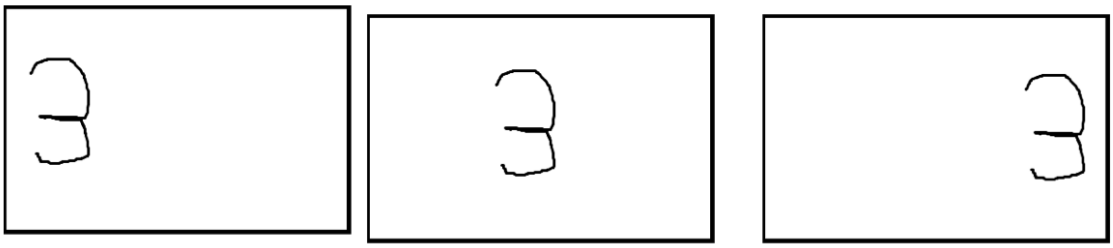

Figure 1: Three spatially shifted instances of the same digit image.

Question 4: Representation Efficiency and Inductive Bias (12 Points)

Modern CNN architectures leverage locality and parameter sharing, whereas fully connected networks treat each input dimension independently.

(a) Compare the number of learnable parameters required to process a $28 \times 28 \times 3$ input using:

(i) a convolutional layer employing three filters of size $2 \times 2 \times 3$ with valid padding, and

(ii) a fully connected layer that reads the same input (i.e., takes all $28 \times 28 \times 3$ pixel values as its input units and maps them to a single hidden layer of your choice).

Explain why one of these approaches is more parameter-efficient, and briefly comment on how this efficiency relates to overfitting.

(b) Again, in reference to Figure 1, explain why weight sharing enables convolutional networks to generalize more robustly to spatial shifts than fully connected networks. Your answer should articulate the inductive bias embedded in convolutional design (for example, the assumption that similar patterns can appear at multiple spatial locations).

Good Luck!