

# Exercise 2: Residual Neural Networks

December 2, 2025

## Questions

- Q1.** You are working as a machine learning engineer at a startup developing an autonomous drone system for forest monitoring. The drone's onboard vision module must process high-resolution imagery in real time, but it has limited computational power. To balance accuracy and efficiency, your team is redesigning one of the convolutional residual blocks used for feature extraction.

The current block consists of a Batch Normalization operation, followed by a ReLU activation and a  $3 \times 3$  convolutional layer. The input and output of the block both contain 128 channels.

- (a) Compute the total number of trainable parameters in the existing block, assuming that both the convolutional and Batch Normalization layers include bias terms.
- (b) To reduce computational cost, a teammate proposes a bottleneck design consisting of three BatchNorm–ReLU–Convolution sequences:
  - A  $1 \times 1$  convolution that reduces the number of channels from 128 to 32
  - A  $3 \times 3$  convolution that keeps the number of channels fixed at 32
  - A  $1 \times 1$  convolution that increases the number of channels back from 32 to 128.

Determine how many trainable parameters this alternative block contains.

- (c) As the lead engineer, you must recommend one design for deployment. Which of the two is more parameter-efficient, and be advantageous for edge devices like drones?
- Q2.** Given SGD with learning rate  $\eta = 0.05$  and weight decay  $\lambda = 0.02$ : If the current weight is  $w = 0.80$  and the gradient from data loss is  $\nabla = 0.30$ , compute the updated weight after one SGD step: [2 points]
- (a) Without weight decay
  - (b) With L2 weight decay (i.e., adding  $\frac{\lambda}{2}\|w\|^2$  to the loss)

## Hint

*The update rules for one SGD step are:*

$$w' = w - \eta \nabla \quad (\text{without weight decay})$$

$$w' = w - \eta(\nabla + \lambda w) \quad (\text{with weight decay})$$

*Alternatively, weight decay can be seen as a multiplicative shrinkage:*

$$w' = (1 - \eta\lambda)w - \eta\nabla$$

## Instructions

- Show all necessary steps clearly.
- Marks are indicated beside each question.