Chapter 9

Convolutional Neural Network

Contents

9.1 How do computers see?

9.2 Computer Vision

9.2.1 Grayscale Model

- Images contain pixels with just one value.
- Can be represented using a 2-D array.
- 0: black, 255: white, 1–254: shades of gray.

9.2.2 RGB Color Model

- Each color channel is stored in 8 bits.
- 8 bits can store 256 values (0–255).
- Also known as 24-bit color (8×3).

9.3 Image Classification

- Can we directly take an image and feed it to a regular fully-connected neural network?
  - Yes, we can, but we will need to first flatten the 2-D image array.
- Issues:

  - No spatial information.
  - Too many parameters.

- Solution:
  - Exploit spatial structure.
  - Each neuron in the hidden layer only respond to a certain set of neurons in the previous layer.
  - Connect the patch in input layer to a single neuron in the subsequent layer.
  - Use a sliding window to define all possible connections.
  - Weighting the connection between the patches and the next layer will allow uss to learn the features.

9.4 Convolutional Neural Network

- CNN or ConvNet is a specialized kind of neural network for processing data that has a known grid-like topology.
  - Image data, which can be thought of as a 2-D grid of pixels.
  - Time-series data, which can be thought of as a 1-D grid taking samples at regular time intervals
- ...
- Convolutional layer performs a transformation called convolution, a specialized king of linear operation on its input.
- In CNN, convolution replaces general matrix multiplication in their convolution layers.
- CNN is specialized for pattern detection.
- Convolutional layer specifies the number of filter kernels each layer must have, and these filters are used to detect patterns.
- Each layer in a convolutional neural network has a 3-D lattice structure.
- Three types of transformations between layers:

  Convolution Apply filters to generate feature maps.

  Activation function To introduce nonlinearity.

  Pooling Downsampling operation on each feature map.

- CNN performs these transformations repeatedly:
  - Higher-order feature detectors after convolution.

  - Lower spatial resolution after pooling.

- In the first stage, the layer performs several convolutions in parallel to produce a set of linear activations.
- In the second stage, each linear activation is run through a nonlinear activation function, such as ReLU. This stage is called the detector stage.
- In the third stage, a pooling function is used to modify the output of the layer further. A pooling function replaces the output of the network at a certain location with a summary statistic of the nearby outputs:
  - The max pooling operation reports the maximum output within a rectangular neighboorhood.
  - Other pooling strategies include average pooling, weighted average pooling, L2 norm, etc.