THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

ECE 599/692 – Deep Learning

Lecture 5 – CNN: The Representative Power

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu

AICIP RESEARCH

Outline

- Lecture 3: Core ideas of CNN
  - Receptive field
  - Pooling
  - Shared weight
  - Derivation of BP in CNN
- Lecture 4: Practical issues
  - The learning slowdown problem
    - Quadratic cost function
    - Cross-entropy + sigmoid
    - Log-likelihood + softmax
  - Overfitting and regularization
    - L2 vs. L1 normalization
    - Dropout
    - Artificial expanding the training set
  - Weight initialization
  - How to choose hyper-parameters
    - Learning rate, early stopping, learning schedule, regularization parameter, mini-batch size,
    - Grid search
  - Others
    - Momentum-based GD
- Lecture 5: The representative power of NN
- Lecture 6: Variants of CNN
  - From LeNet to AlexNet to GoogleNet to VGG to ResNet
- Lecture 7: Implementation
- Lecture 8: Applications of CNN

AICIP RESEARCH

The universality theorem

- Neural networks with a single hidden layer can be used to approximate any continuous functions to any desired precision

Visual proof

AICIP RESEARCH

- One input and one hidden layer
  - Weight selection (first layer) and the step function
  - Bias selection and the location of the step function
  - Weight selection (2nd layer) and the rectangular function (”bump”)
- Two inputs and two hidden layers
  - From “bump” to “tower”
- Accumulating the ”bumps” or “towers”

THE UNIVERSITY OF TENNESSEE KNOXVILLE

4

Beyond sigmoid neuron

AICIP RESEARCH

- The activation function needs to be well defined as z goes to both positive and negative infinity
- What about ReLU?
- What about linear neuron?

THE UNIVERSITY OF TENNESSEE KNOXVILLE

5

Why deep network?

AICIP RESEARCH

- If two hidden layers can compute any function, why multiple layers or deep networks?
- Shallow networks require exponentially more elements to compute than do deep networks

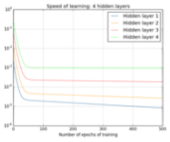
THE UNIVERSITY OF TENNESSEE KNOXVILLE

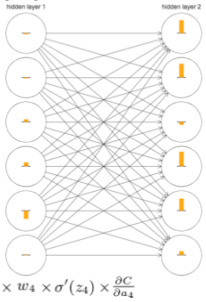
6

AICIP RESEARCH

Why are deep networks hard to train?

- The unstable gradient problem
  - Gradient vanishing
  - Gradient exploding





$$\frac{\partial C}{\partial b_1} = \sigma'(z_1) \times w_2 \times \sigma'(z_2) \times w_3 \times \sigma'(z_3) \times w_4 \times \sigma'(z_4) \times \frac{\partial C}{\partial a_4}$$

AICIP RESEARCH

Acknowledgement

- All figures from this presentation are based on Nielsen's NN book, Chapters 4 and 5.