THE UNIVERSITY OF TENNESSEE KNOXVILLE

AICIP RESEARCH

ECE 599/692 – Deep Learning

Lecture 4 – CNN: Practical Issues

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
  - Normalized input and initialization of hyperparameters
  - Cross validation
  - Momentum
  - Learning rate
  - Activation functions
  - Pooling strategies
  - Regularization
- Lecture 5: Variants of CNN
  - From LeNet to AlexNet to GoogleNet to VGG to ResNet
- Lecture 6: Implementation
- Lecture 7: Applications of CNN – Binary hashing
- Lecture 8: Applications of CNN – Person re-identification

AICIP RESEARCH

Cost functions

- Least square quadratic
  - The learning slowdown problem
- Sigmoid and cross-entropy cost
  - How does it solve the slowdown problem?
- Softmax and log-likelihood cost

Activation functions

AICIP RESEARCH

- Sigmoid
- Softmax
- Tanh
- ReLU
- LeakyReLU

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Regularization methods

AICIP RESEARCH

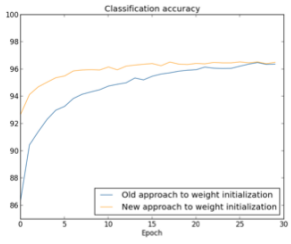
- The problem of overfitting
- Weight decay
  - L^2 normalization
  - L^1 normalization
- Dropout
- Artificial expansion of the training data

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Weight initialization

AICIP RESEARCH

- Gaussian distribution with different std



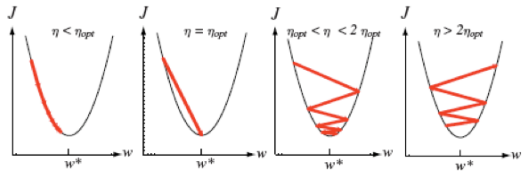
THE UNIVERSITY OF TENNESSEE KNOXVILLE

How to choose hyper-parameters?

- The learning rate η
- Early stopping
- Learning rate schedules
- Regularization parameter λ
- Mini-batch size m
- Grid search and the automated technique

Learning rates

- Always use smaller rates



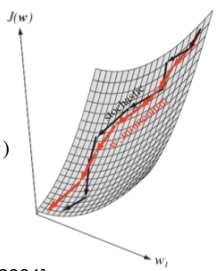
From [Duda&Hart:2001]

Momentum

- Taking into account of previous changes

$$\omega_{st}^{k+1} = \omega_{st}^{k} - c^{k}\frac{\partial E^{k}}{\partial \omega_{st}^{k}}$$

$$\omega_{st}^{k+1} = \omega_{st}^{k} + (1-c^{k})\Delta\omega_{bp}^{k} + c^{k}(\omega_{st}^{k} - \omega_{st}^{k-1})$$



From [Duda&Hart:2001]