Default Baseline Models

Sargur N. Srihari

srihari@cedar.buffalo.edu

Topics

1. Performance Metrics
2. Default Baseline Models
3. Determining whether to gather more data
4. Selecting hyperparamaters
5. Debugging strategies
6. Example: multi-digit number recognition

Baseline Approaches

- After choosing performance metrics and goals, next step is to establish an end-to-end system

- Here we provide recommendations for which algorithms to use as first baseline approach

- May want to begin without using deep learning

  – There may be a chance of choosing a few linear weights as in logistic regression

  – If problem is AI-complete such as object recognition, speech recognition , machine translation, etc appropriate deep learning model can be used

Architecture depends on input data

1. Supervised learning with fixed size vector input: use a feedforward network with fully connected layers

2. Input has a topological structure, e.g., images: use a CNN

    – Use a piecewise ReLU (Leaky ReLU, PreLU or maxout)

3. Input or output is a sequence: use a gated recurrent net (LSTM or GRU)

Choice of Optimization Algorithm

- SGD with momentum with a decaying learning rate
  - Linear decay
  - Exponential decay
  - Decreasing learning rate by factor of 2-10 when validation error rate plateaus

- Adam batch normalization can have a dramatic effect on optimization performance

- Reasonable to omit batch normalization from very first baseline unless optimization is problematic

5

Regularization

- Unless training set includes tens of millions of examples, should include some form of regularization from start

- Early stopping should used universally

- Dropout is easy to implement amd compatible with many models/training