Large Scale Deep Learning and Self-Supervision in Vision and NLP

Thomas Breuel

NVIDIA

INTRODUCTION

Topics

- using large amounts of unlabeled training data

- self-supervised training

- LSTM, transformers, ViT

- large language models

- diffusion models

Application: OCR (text recognition, layout recognition, information extraction)

Different Kinds of Learning

- supervised learning
  - inputs and outputs are given

- unsupervised learning
  - only inputs are given

- semi-supervised learning
  - combine supervised and unsupervised data

- self-supervised learning
  - algorithm derives a supervised problem from unsupervised data

Different Kinds of Learning

- transfer learning
  - use a model trained on one problem to solve another problem
- active learning
  - the learning algorithm requests transcribed data from an oracle
- metric learning
  - learn a distance measure that helps with clustering / classification
- representation learning
  - transform input vectors into another space that, removes noise, expresses invariances, places "similar" samples closer together

Different Kinds of Models

- discriminative vs generative

- linear models

- non-linear models
    - parametric probabilistic
    - support vector machines
    - deep learning
        - convolutional
        - recurrent
        - transformer

Overview

Motivation:

- OCR example

Classical Techniques:

- statistical basis of machine learning (mostly review)

- linear methods (mostly review?)

Deep Learning:

- language modeling and sequence learning

- self-supervised learning for images

- generative modeling (VAE, flows, GANs)

Background

I'm assuming you have

- done some supervised deep learning, e.g., trained MNIST, ImageNet, etc.
- understand the basics of SGD, loss functions, etc.

Focus

We will focus on the unsupervised / self-supervised aspects of methods and papers we will be discussing, discussing other aspects of papers only as needed.