Deep Learning Review

- **machine learning**
  - conventional ml
    - were limited in their ability to process natural data in their raw form
    - required careful engineering and considerable domain expertise to design features
  - representation learning
    - automatically discover the representations from raw data
    - deep learning
      - multiple levels of representation
      - transform the representation at one level into a representation at a higher, slightly more abstract level
      - good at discovering intricate structures in high-dimensional data
      - requires very little engineering by hand
      - take advantage of increases in the amount of available computation and data
- **supervised learning**
  - basic
    - labelled examples
    - adjustable weights
    - objective function that measures the error
  - learning algorithm
    - computes a gradient vector
    - the negative gradient vector indicates the direction of steepest descent
    - stochastic gradient descent (SGD)
      - small set of examples
  - generalization ability
    - its ability to produce sensible answers on new inputs that it has never seen during training
  - linear classifiers
    - only carve their input space into very simple regions, namely half-spaces separated by a hyperplane
    - can not be insensitive to irrelevant variations of the input
    - hand design good feature extractors, which requires a considerable amount of engineering skill and domain expertise
  - deep learning
    - good features can be learned automatically
    - multilayer stack of simple modules
    - can implement extremely intricate functions
      - sensitive to minute details
      - insensitive to large irrelevant variations
- **backpropagation**
  - train multilayer architectures
    - stochastic gradient descent
    - compute gradients using the backpropagation procedure
      - a practical application of the chain rule of derivatives
  - late 1990s
    - forsaken by the machine-learning community and ignored by the computer-vision and speech-recognition communities
    - issues
      - poor local minima
      - saddle points
  - revived around 2006
    - introduced unsupervised learning procedures that could create layers of feature detectors without requiring labelled data
  - 2012
    - versions of the deep net from 2009 were being developed by many of the major speech groups and were already being deployed in Android phones
- **convolutional neural networks**
  - process data that come in the form of multiple arrays
  - key ideas
    - local connections
    - shared weights
    - pooling
    - many layers
  - stages
    - first
      - convolutional layers
        - to detect local conjunctions of features from the previous layer
      - pooling layers
        - to merge semantically similar features into one
    - second
      - two or three stages of convolution, non-linearity and pooling are stacked
      - followed by more convolutional and fully-connected layers
- **distributed representations and language processing**
  - two exponential advantages of deep nets that use distributed representations
    - enable generalization to new combinations of the values of learned features beyond those seen during training
    - composing layers of representation in a deep net brings the potential for another exponential advantage
  - word vector
    - predict the next word in a sequence from a local context of earlier words
    - each word in the context is presented to the network as a one-of-N vector
    - in the first layer, each word creates a different pattern of activations, or word vectors
- **recurrent neural networks**
  - introduce
    - sequential inputs
    - one element at a time
    - state vector
      - historical information
  - issue
    - gradients explode or vanish
      - grow or shrink at each time step
    - solved by advances in architecture and ways of training
  - usage
    - predict next character or word
    - machine translation
      - encoder
        - generate state vector of the thought expressed by the sentence
      - decoder
        - state vector from encoder as input
        - outputs probability distribution of the words
    - image to word
      - encoder
        - CNN
      - decoder
        - RNN
  - augment RNNs
    - conventional RNNs
      - it is difficult to learn to store information for very long
    - LSTM
      - memory cell
      - machine translation
    - Neural Turing Machine
      - 'tape-like' memory
        - read from or write to
      - be taught 'algorithms'
        - output a sorted list of symbol
    - memory networks
      - associative memory
      - question-answering
      - answering questions that require complex inference
    - *reasoning and symbol manipulation*
- **the future of deep learning**
  - unsupervised learning
  - combining deep learning and reinforcement learning
  - natural language understanding
  - artificial intelligence come about

XMind:ZEN | 试用模式