Formal Systems

Cognition, cognitive theory are formal systems, we use mathematics to study them

Abstract concepts can be reasoned about precisely when situated in formal systems

Neural networks are continuous systems

Constructing the continum

Axiomatization

> Describe the basic properties and declare them to be true by definition

Construction

> Use simpler objects and operations to explicitly define more complex models

Equivalence classes

> Partitions a set based on some rules

Dynamic Systems

Recurrent ,learning, and biological neural networks

Motor control is the effector of a dynamic system

The mind is an abstract dynamical system with continuous state variables that are not activation values of units or representations

NNs as Probabilistic Models

Used for	stochastically searching for global optima
	representing and rationally coping with uncertainty
	measuring information
Deployed in	neural network models
	symbolic modes with non-determinism and uncertainty (e.g. inferring knowledge from experience, using knowledge to infer outputs given inputs)

Optimization

Processing

> Activation dynamics. Maximizes well-formedness (harmony) of the activation patter (depends on the connection weight). Spreading activation dynamics is an optimization algorithm for the representation.

Learning

> Weight dynamics. Minimizes error. Weight-adjustment dynamics is an optimization algorithm for the knownledge in the weights: learning algorithm

Probabilistic modelling

> Parameters of the statistical model change as more data is received. Optimized based of likelihood according to data or Bayesian posterior probability of the data

Fourier analysis

f(x) = \sum_k c_k e^{ikx} employs a basis of imaginary powers of x, {e^{ikx}}_{k \in \Z}

Also a basis of cos(kx) and sin(kx)

Fourier coefficient states how strongly an oscillation of frequency 1/k is present in f

{f(t)}_t describe f in the time / spatial domain

{c_k}_k describe f in the frequency / spatial-frequency domain

Support Vector Machines

Use supervised learning to learn a region of activation space for each concept

Classification driven only by training near the region boundary

Wide margin: error function favors a large margin between the training samples and the boundary it posits for separating the categories

Support Vector Machines (cont)

Slack variable: minimizes a variable for each training example that "picks up the slack" between the point and the category region it should be in

Kernel trick: implicitly maps the data into a high dimensional space in which classification conceptually takes places (implemented through a kernel function)

Discrete structures of distribution patterns

vectors v in R: distributed representation

With respect to an appropriate conceptual basis for V, components of a representation v indicate the strength of a set of basis concepts in v: gradient conceptual representation

Eigen-basis re-scales components

Analyse the entire distribution within V of the representations {v^k} of a set of represented items {x^k}

Clusters of {v^k} constitute a conceptual group and may be hierachically structured

Can construct is such that greater distance between v^{k} and v^{t} means greater mental distinguishability of x^{k} and x^\{t}

Harmony

Weight matrices, error functions, learning as optimization

Activation vectors, well-formedness = harmony function, processing as optimzation

- Parallel, violable-constraint satisfaction

- Schemas/prototypes in Harmony landscapes

Local optima is a deterministic problem, global optima require randomized/stochastic algorithms

By **goldmist**
cheatography.com/goldmist/

Published 9th December, 2016.
Last updated 9th December, 2016.

Inductive learning

Finding the best hypothesis within a hypothesis space about the word that the learning is trying to understand

Goodness of a hypothesis is determined jointly by how well H fits the data that the learning has received about the world and how simple H is

A hypothesis is a probabilistic data-generator

Maximum Likelihood Principle

Maximum A Posteriori Principle: Bayesian principle that says pick the hypothesis that has highest a posteriori probability, balancing the likelihood of the data against the a priori probability of H. Best not pick a H, maintain a degree of belief for every H in the H space

Maximum Entropy Principle: Maxent. Pick the H with the max missing information, among those H that are consistent with the known data

Minimum Description Length Principle: shorter is better

By **goldmist**
cheatography.com/goldmist/

Published 9th December, 2016.
Last updated 9th December, 2016.