Machine Learning

Martin Sewell
Department of Computer Science
University College London

April 2007 (last updated January 2009)

1 Introduction

Machine learning is an area of artificial intelligence concerned with the study of computer algorithms that improve automatically through experience. In practice, this involves creating programs that optimize a performance criterion through the analysis of data. Machine learning can be viewed as an attempt to automate 'doing science'. For introductory texts, see Langley (1996), Mitchell (1997), Alpaydin (2004) and Bishop (2006). Mitchell has long been considered the 'bible', but is now slightly dated. Bishop (2006) is the first textbook on pattern recognition to present the Bayesian viewpoint. For introductory books on computational learning theory (which emphasizes the 'probably approximately correct' (PAC) model of learning (Valiant 1984)), see Anthony and Biggs (1992) and Kearns and Vazirani (1994). As the following taxonomy shows, machine learning algorithms may be categorised in at least six ways.

2 Taxonomy

- Model type

  probabilistic Build a full or partial probability model.

  non-probabilistic Find a discriminant/regession function directly.

- Type of reasoning

  Induction Reasoning from observed training cases to general rules, which are then applied to the test cases.

  Transduction Reasoning from observed, specific (training) cases to specific (test) cases. Figure 1 (page 2) shows the relationship between induction and transduction.

- Type of machine learning

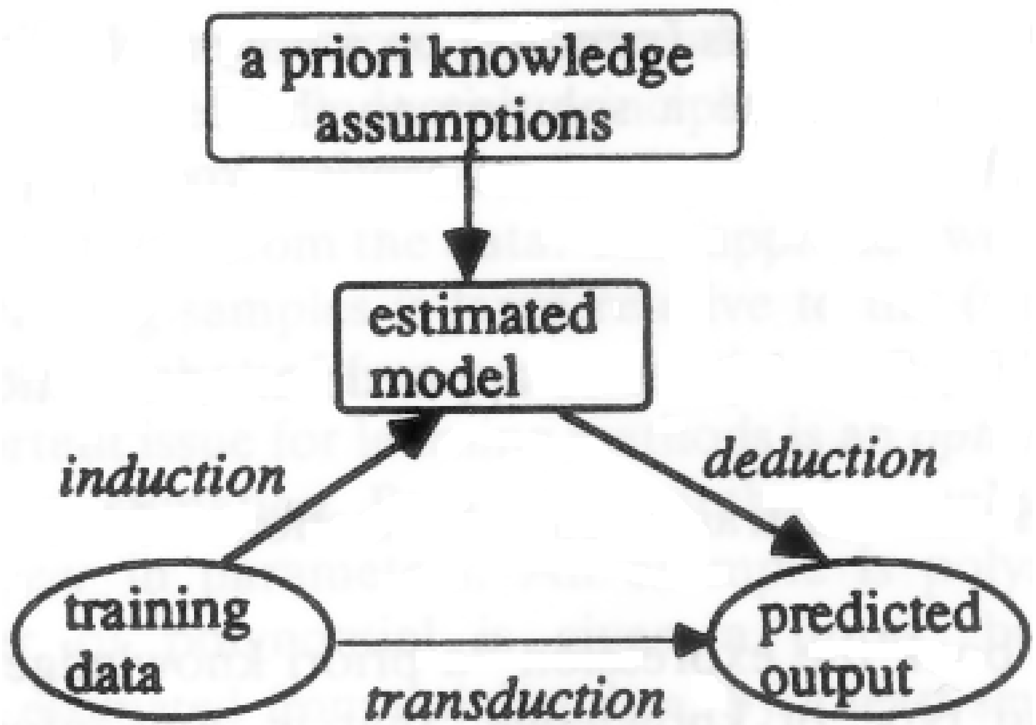

Figure 1: Two types of inference: induction-deduction and transduction (Cherkassky and Mulier 1998)

Supervised learning The algorithm is first presented with training data which consists of examples which include both the inputs and the desired outputs, thus enabling it to learn a function. The learner should then be able to generalize from the presented data to unseen examples. In situations where there is a cost to labeling data, a method known as *active learning* may be used, where the learner chooses which data to label.

Unsupervised learning The algorithm is presented with examples from the input space only and a model is fit to these observations. For example, a clustering algorithm would be a form of unsupervised learning.

Reinforcement learning An agent explores an environment and at the end receives a *reward*, which may be either positive or negative. In effect, the agent is told whether he was right or wrong, but is not told how. Examples include playing a game of chess (you don't know whether you've won or lost until the very end) or a waitress in a restaurant (she has to wait for the end of the meal before she finds out whether or not she receives a tip).

- The manner in which the training data are presented to the learner

Batch All the data are given to the learner at the start of learning.

On-line The learner receives one example at a time, and gives an estimate of the output, before receiving the correct value. The learner updates its current hypothesis in response to each new example and the quality of learning is assessed by the total number of mistakes made during learning.

- Task

Classification May be *binary* or *multi-class.*

Regression Real-valued targets (generalizes classification).

- Classification model type

generative model Defines the joint probability of the data and latent variables of interest, and therefore explicitly states how the observations are assumed to have been generated.

discriminative model Focuses only on discriminating one class from another.

3 Desirable Features

Desirable features of a machine learning algorithm:

- *Simple* solutions are appropriately favoured over complicated ones.
- *Powerful* enough to learn the solution to a given problem.
- *Stable* to parameter variations.
- *Converges* in finite time.
- *Scales* reasonably with the number of training examples, the number of input features and the number of test examples.

References

ALPAYDIN, Ethem, 2004. *Introduction to Machine Learning.* Adaptive Computation and Machine Learning. Cambridge, MA: The MIT Press.

ANTHONY, Martin, and Norman BIGGS, 1992. *Computational Learning Theory.* Volume 30 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge: Cambridge University Press.

BISHOP, Christopher M., 2006. *Pattern Recognition and Machine Learning.* Information Science and Statistics. New York: Springer.

CHERKASSKY, Vladimir, and Filip MULIER, 1998. *Learning from Data: Concepts, Theory, and Methods.* Adaptive and Learning Systems for Signal Processing, Communications and Control Series. New York: Wiley.

KEARNS, Michael J., and Umesh V. VAZIRANI, 1994. *An Introduction to Computational Learning Theory.* Cambridge, MA: The MIT Press.

LANGLEY, Pat, 1996. *Elements of Machine Learning.* San Francisco, CA: Morgan Kaufmann.

MITCHELL, Tom M., 1997. *Machine Learning.* New York: McGraw-Hill.

VALIANT, L. G., 1984. A Theory of the Learnable. *Communications of the ACM*, **27**(11), 1134–1142.