18 LEARNING FROM EXAMPLES

# 18.1 Forms of Learning

Any component of an agent can be improved by learning from data. The improvements, and the techniques used to make them, depend on four major factors:

* Which component is to be improved.
* What prior knowledge the agent already has.
* What representation is used for the data and the component.
* What feedback is available to learn from.

**Components to be learned**

**Representation and prior knowledge**

**Feedback to learn from**

There are three types of feedback that determine the three main types of learning:

* unsupervised learning the agent learns patterns in the input even though no explicit feedback is supplied
* reinforcement learning the agent learns from a series of reinforcements—rewards or punishments.
* supervised learning the agent observes some example input—output pairs and learns a function that maps from input to output.

In semi-supervised learning we are given a few labeled examples and must make what we can of a large collection of unlabeled examples. Even the labels themselves may not be the oracular truths that we hope for.

# 18.2 supervised Learning

Học có giám sát là chúng ta huấn luyện mô hình máy học bằng một tập dữ liệu chứa input and output mong muốn

Khi output nằm trong tập hữu hạn các giá trị thì chúng ta gọi the learning problem là Classification (phân loại). Và sẽ được gọi là Boolean or binary classification if there are only two values.

Khi output nó là 1 number thì the learning problem is called regression

# 18.3 Learning decision tree

**The decision tree representation**

A decision tree represents a function that takes as input a vector of attribute values and returns a "decision"—a single output value.

A decision tree reaches its decision by performing a sequence of tests.

**Expressiveness of decision trees**

A Booleandecision tree is logically equivalent to the assertion that the goal attribute is true if and only if the input attributes satisfy one of the paths leading to a leaf with value true.

**Inducing decision trees from examples**

The DECISION-TREE-LEARNING algorithm adopts a greedy divide-and-conquer strategy: always test the most important attribute first.

**Choosing attribute tests**

Để chọn thuộc tính test. Chúng ta tính Entropy của từng thuộc tính (entropy càng cao thì càng quan trọng)

Entropy( the notion of information gain ) : is a measure of the uncertainty of a random variable; acquisition of information corresponds to a reduction in entropy.

**Generalization and overfitting**

**Overfitting** is an undesirable machine learning behavior that occurs when the machine learning model gives accurate predictions for training data but not for new data. (nghĩa là nó hoạt động tốt trên dữ liệu đào tạo nhưng kém hiệu quả trên tập dữ liệu mới)

**decision tree pruning** combat với overfitting. Pruning work by eliminating nodes that are not clearly relevant.

**early stopping**(là một kết hợp của phương pháp pruning và information gain): the decision tree algorithm stop generating nodes when there is no good attribute to split on, rather than going to all the trouble of generating nodes and then pruning them away.

**Một vài cái problem của desicion tree**:

* Missing data
* Multivalued attrubutes
* Continuous and integer-valued input attributes
* Continuous-valued output attributes:

# 18.4 Evaluating and choosing the best hypothesis

# 18.5 the theory of Learning

# 18.6 regresstion and Classification with linear Models

# 18.7 Artifitial Neural Networks

# 18.8 Nonparametric Models

# 18.9 Support Vector Machines

# 18.10 Ensemble Learning

# 18.11 practical Machine learning

# 18.12 Summary