

Machine Learning Study Guide (AIMA 19.1-19.6)

Artificial Intelligence

1 Machine Learning

According to Tom Mitchell, machine learning is the study of algorithms that

- improve their performance P
- at some task T
- with experience E .

A well-defined learning task is given by $\langle P, T, E \rangle$.

1. Formulate the following problems according to Tom Mitchell's machine learning problem specification (see [Machine Learning Slides](#)) and the specification our textbook. For each of the following problems specify:

- The task T ,
- The performance measure P ,
- The experience E ,
- The target function $f : \mathcal{X} \rightarrow \mathcal{Y}$, that is,
 - the input space \mathcal{X} , and
 - the output space \mathcal{Y} .

Remember that a function maps a domain to a co-domain, and these domains are sets.

(a) Medical diagnosis: A patient walks in with a medical history and some symptoms, and you want to identify the problem.

Solution:

- Task, T : diagnose problem
- Performance, P : diagnosis is correct or incorrect
- Experience, E : $\langle \text{medical} - \text{history}, \text{symptoms} \rangle$
- Target function $f : \mathcal{X} \rightarrow \mathcal{Y}$:
 - $\mathcal{X} = \{\vec{x} | x_1 \in \{\text{family} - \text{history} - \text{heart} - \text{disease}\}, x_2 \in \mathbb{R} = \text{cholesterol} - \text{level}\}$
and other such features
 - $\mathcal{Y} = \{\text{disease}_1, \text{disease}_2, \dots, \text{disease}_n\}$

2. Approximating a function $\mathbf{X} \mapsto \mathbf{y}$ given a set of instances \mathbf{X} and associated labels \mathbf{y} is an example of ...

Solution: supervised learning.

3. Predicting the price of a house given its characteristics, e.g., number of rooms, square footage, etc., is an example of ...

Solution: regression.

4. Predicting the object contained in an image from a finite list of options is an example of ...

Solution: classification.

2 Supervised Statistical Machine Learning

1. In a supervised learning problem, you are given a labeled data set with which to develop and train a classifier. What is the most important step in the process before choosing a model/hypothesis class and training the classifier?

Solution: Set aside a test set, and optionally a validation set, and only use the test set for testing the model you develop and training.

2. What does *training* mean in the context of statistical machine learning?

Solution: Iteratively adjusting the learnable parameters of a statistical model to minimize some error function.

3. What is a *hyperparameter*?

Solution: A hyperparameter is a parameter to the learning algorithm or fixed parameters of the model, as opposed to the parameters that are adjusted, aka *learned* by the algorithm in the training process. Hyperparameters are fixed at the beginning of training and affect the training process in various ways. Examples of hyperparameters include learning rate, neural network architecture (layers and nodes), and choices of neural network activation functions.

4. Distinguish between *tabular data* and *unstructured data*.

Solution: Tabular data is also called *structured data*. We say it is structured because we impose the structure on it. There is nothing inherent in the data that requires a particular feature to come before another, but we must choose some order and stick with it.

Unstructured data is data whose structure is inherent in the data, not imposed by us. Examples include images and natural language text.