Notes on Foundations of Machine Learning by M. Mohri

compiled by D. Gueorguiev 3/16/2025

# The Probably Approximately Correct (PAC) Learning Model

- the set of all possible *examples* or *instances*. also is referred to as the *input space*.

– the set of all possible labels / target values. When we have *binary classification*.

A *concept* is a mapping from to . Since , we can identify with the subset of over which it takes the value 1.

A *concept class* is a set of concepts.

Let us assume that the examples are independently and identically distributed (i.i.d.) according to some fixed but unknown distribution . The learner considers a fixed set of possible concepts called hypothesis set, which may not coincide with . The learner receives a sample drawn i.i.d. according to as well as the labels set which are based on a specific target concept to learn. The learner’s task is to use the labeled sample to select a hypothesis that has a small *generalization error* with respect to the concept . The generalization error of a hypothesis , also referred to as the *true error* of is denoted with and is defined as follows:

**Definition** *Generalization error*

Given a hypothesis , a target concept and an underlying distribution , the generalization error or risk of is defined by

(1)

where is the indicator function of the event .

The generalization error of a hypothesis is not directly accessible to the learner since both the distribution and the target concept are unknown. However, the learner can measure the *empirical error* of a hypothesis on the labeled sample .

Definition Empirical error

# References

[1] [Foundations of Machine Learning, Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar, 2012](https://github.com/dimitarpg13/statistical_learning_and_kernel_methods/blob/main/literature/books/Foundations_of_Machine_Learning_Mohri_2012.pdf)

[2]

# Appendix

## Cross-validation

in practice the amount of labeled data available is often too small to set aside a validation sample since that would leave an insufficient amount of training data. Instead, a widely adopted method known as n-fold cross-validation is used to exploit the labeled data both for model selection (selection of the free parameters of the algorithm) and for training.

Let denote the vector of free parameters of the algorithm. For a fixed value of , the method consists of first randomly partitioning a given sample of labeled examples into subsamples or *folds*. The -th fold is thus a labeled sample of size . Then, for any , the learning algorithm is trained on all but the th fold to generate a hypothesis , and the performance of is tested on the -th fold.