

NOTATIONS

In this note we review notational conventions used throughout the class. Sometimes we use different notations for the same object to emphasize different aspects of it.

- **Sample sequence:** $S^m = s_1, s_2, \dots, s_m$. Sometimes we shall drop the m and write S instead of S^m . In the supervised learning setup $s_i = (x_i, y_i)$ where x_i is the domain variable and where y_i denotes the label.
- **Probability:** $P(X \in \mathcal{A})$ means “probability that X belongs to set \mathcal{A} when X is distributed according to P ”. Alternatively we use $Pr_{X \sim P}(X \in \mathcal{A})$ or $\mathbb{P}_{X \sim P}(X \in \mathcal{A})$. Similarly, we write $P^m(S^m \in \mathcal{B})$ whenever the S_i ’s are i.i.d. $\sim P$
- **Loss functions:**
 - $L_S(h)$ refers to the empirical risk (or training error) with respect to S . For instance, for the 0-1 loss function it is equal $(1/m)|\{i : h(x_i) \neq y_i\}|$.
 - $L_{P,f}(h)$ refers to test error, or the generalization error, and is the loss with respect to distribution P and function h when the true function (to be learned) is f . For instance, for the 0-1 loss function $L_{P,f}(h) = P(f(X) \neq h(X))$.
- **Algorithm:** given S the learner produces an algorithm $A(S)$ which defines a predictive rule $\hat{f}_S(X)$. In other words, we have $S \rightarrow A(S) \rightarrow \hat{f}_S(X)$. For example, in HW1 ex3, $A(S)$ is the algorithm that takes S and produces the (rectangular) function $\hat{f}_S(X)$. The empirical loss with respect to $\hat{f}_S(X)$ is thus $L_S(\hat{f}_S)$. Abusing somewhat notation, sometimes we shall write $L_S(A(S))$ instead of $L_S(\hat{f}_S)$.
- **Empirical Risk Minimization (ERM):** given a set of functions \mathcal{H} and a sample sequence S the ERM rule selects the function that minimizes the empirical loss, that is

$$\text{ERM}_{\mathcal{H},S} = \arg \min_{h \in \mathcal{H}} L_S(h).$$