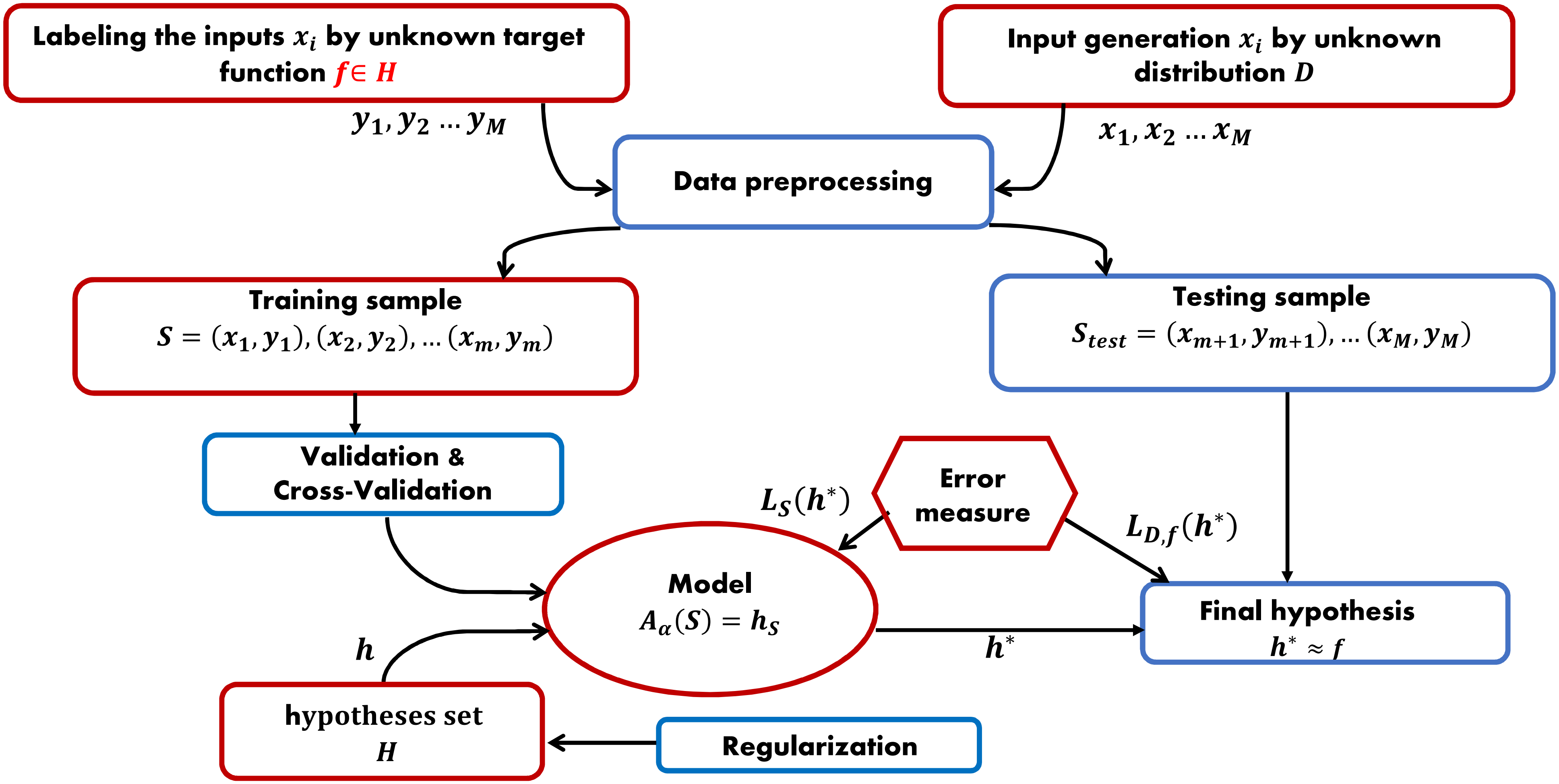
Part 1: Machine learning theory

1. **Learning framework:**
   1. **ERM algorithm.**
   2. **PAC Learning model.**
   3. **SLPOA : General Learning model**
2. **Uniform convergence**
3. **Learnability of infinite size hypotheses classes**
4. **Tradeoff Bias/Variance**
5. **Non-Uniform learning.**

Supervised Learning Passive Offline Algorithm (SLPOA)



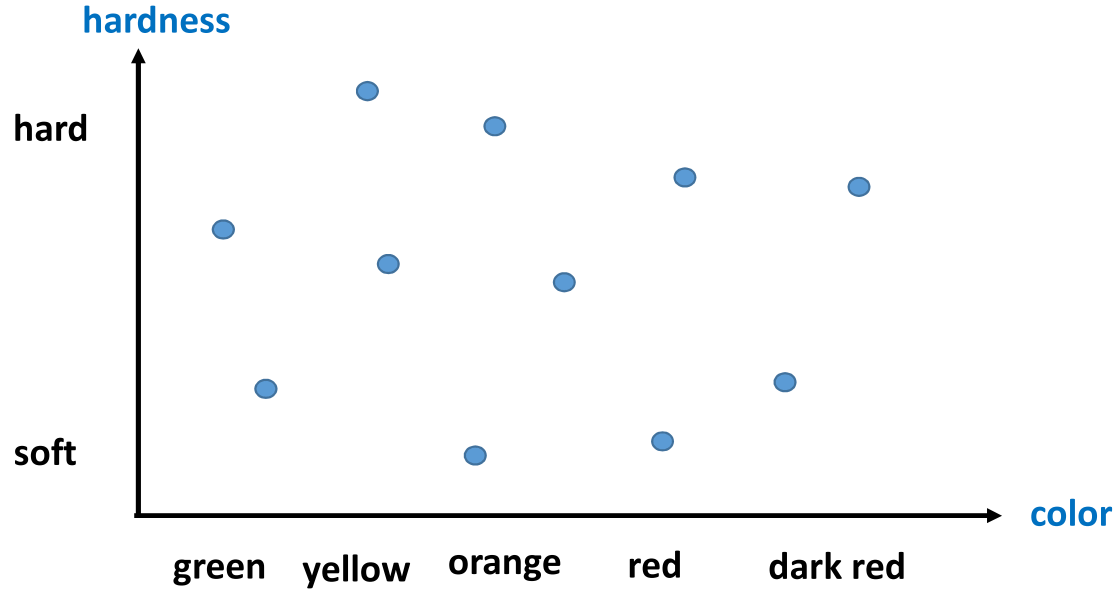
1.Learning framework

Example: Predict the taste (non-delicious, delicious) of tomatoes found in the market.

The features of tomatoes $x_i = (Color, Hardness)$:

- Color: {green, yellow, orange, red, dark red}.

- Hardness: {hard, soft}.



1.Learning framework

Objective of learning:

Given S, find the hypothesis $\boldsymbol{h}^* = arg\min_{h} \boldsymbol{L_S}(\boldsymbol{h})$ having the smaller error.

Inputs of the learning algorithm:

- **Training set:** $S = \{(x_1, y_1), \dots, (x_m, y_m)\} \in (X \times Y)^m$, such that the points x_i are sampled $(i.i.d.)$ by a probability distribution $\mathcal{D}$ on X and labeled by a target function $f\colon X \rightarrow Y$.
- **Features set:** $X = [0,1]^2$
- **Labels set:** $Y = \{Delicious, non - Delicious\} = \{0,1\}$

Outputs of the learning algorithm:

- **Optimal hypothesis:** $A_\alpha(S) = h_s$ such that: $A_\alpha\colon \bigcup_{m=1}^{\infty} (X \times \{0,1\})^m \longrightarrow \{h, h\colon X \longrightarrow \{0,1\}\}$
- **Error measure:** (general error)

$$L_{\mathcal{D},f}(h) \stackrel{\text{def}}{=} \underset{x \sim \mathcal{D}}{\mathrm{P}}[h_s(x) \neq f(x)] \stackrel{\text{def}}{=} \mathcal{D}(\{x\colon h_s(x) \neq f(x)\})$$

1.Learning framework: Interpretation

- Training step
  - $A_\alpha: \bigcup_{m=1}^{\infty}(X \times \{0,1\})^m \longrightarrow \{h, h: X \longrightarrow \{0,1\}\}$
  - $S \in \bigcup_{m=1}^{\infty}(X \times \{0,1\})^m \longrightarrow A_\alpha(S) = h_s \in \{h, h: X \longrightarrow \{0,1\}\}$
  - If $|S| = m$ then $S \in (X \times \{0,1\})^m$
  - $\{h, h: X \longrightarrow \{0,1\}\} \rightarrow h(x) = y \in \{0,1\}$
  - $A_\alpha(S) = h_s = \mathbf{argmin}_h \boldsymbol{L_S(h)} \rightarrow L_S(h_s) \in [0,1]$ if $L_S(h_s) \approx 0$ then we have a good approximation

- Testing step

$$L_{\mathcal{D},f}(h_s) \stackrel{\text{def}}{=} \underset{x \sim \mathcal{D}}{\mathrm{P}}[x \in X:\ h_s(x) \neq f(x)] \stackrel{\text{def}}{=} \mathcal{D}(\{x: h_s(x) \neq f(x)\}) \in [0,1]$$

$L_{\mathcal{D},f}(h_s) \approx 0$ then we have a good generalization : there is a learning

1.1. ERM algorithm

The learning algorithm ERM (Empirical Risk Minimization) aims to select h_S :

$$\boldsymbol{ERM(S) = h_S \in \underset{h}{\operatorname{argmin}}\{L_S(h)\}}$$

Such that:

$$L_s(h) \stackrel{\text{def}}{=} \frac{|\{i \in I : h(x_i) \neq y_i\}|}{|S|} \qquad I = \{1,2,\dots,m\}$$

Notice:

Given the following hypothesis:

$$\boldsymbol{\bar{h}_S(x) = \begin{cases} y_i & if\ \exists i \in I\,, x_i = x\,, \qquad (if\ x \in S) \\ 0 & otherwise\ \ (if\ x \notin S\ then\ x \in S_{test}) \end{cases}}$$

we have: $\boldsymbol{L_s(\bar{h}_S) = \frac{1}{m}\sum_{i=1}^{m} 1_{[h(x_i) \neq y_i]} = 0}$.

So, with a high probability, this hypothesis can be chosen by the algorithm ERM.

$$\boldsymbol{L_{D,f}(h_S) \approx L_{test}(h_S) = \frac{\sum_{i=m+1}^{M} \mathbb{1}_{[y_i \neq h_S(x_i)]}}{M-m} = 1 \gg L_s(\bar{h}_S) = 0}$$

Then we have an overfitting, bad generalization

1.1. ERM_H algorithm

The learning algorithm ERM_H aims to apply ERM to a set of limited hypothesis search, called set of hypotheses H, such that:

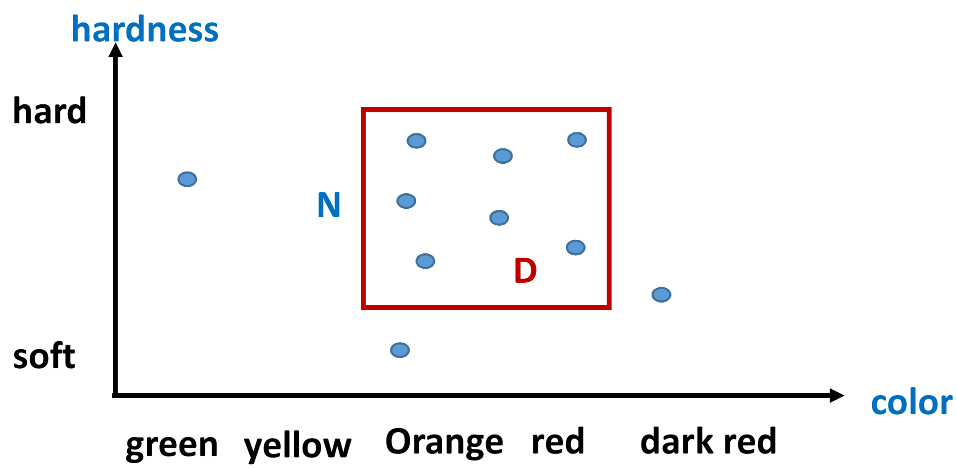
$$H \subseteq \{h:\ X\ \rightarrow\ Y\} \qquad ERM_H(S) = h_S \in \underset{h \in H}{\operatorname{argmin}}\{L_S(h)\}$$

The choice of *H*: H must be chosen based on prior knowledge, before ERM_H sees the data.

Preliminary knowledge: there is a hypothesis in the form of a rectangle, such that inside the rectangle the tomatoes are delicious and outside they are not delicious.

$\boldsymbol{H = \{w = (l, L): l, L \in [0,1]^2\} \rightarrow |H| \approx \infty}$**:**

collection of rectangles.



1.1. ERM_H algorithm

Definition: realizability hypothsesis

There exist a hypothesis $h^* \in H$ such that $L_{\mathcal{D},f}(h^*) = 0$

Lemma:

If realizability hypothesis is respected Then :

- with probability 1 we have :

$$\forall S \subset X \quad L_S(h^*) = 0$$

- If we use ERM_H to look for the best hypothesis h_S, then with probability 1 we have:

$$L_S(h_S) = 0$$

1.1. ERM_H algorithm

Theorem: Generalization bound of learning

- Let
  - X be the set of inputs, $Y = \{0,1\}$ set of labels, and $\boldsymbol{H}$: **finite hypothesis set**.
  - $S = \{(x_i, y_i), i \in I\}$, $S_x = \{x_i, i \in I\}$ and $I = \{1, \ldots, m\}$
- Assume that
  - S is generated **(i. i. d.)** by an unknown probability distribution $\mathcal{D}$ on X $(\mathbf{S_x} \sim \boldsymbol{\mathcal{D}^m})$ and labeled by an unknown target f: $X \rightarrow Y$ such that f(x_i)=y_i $\forall i \in I$, where the **realizability hypothesis is respected**.

So, $\forall \boldsymbol{\varepsilon}, \exists \boldsymbol{\delta}$ such that ERM_H is able to generate a hypothesis $h_S \in \underset{h \in H}{\mathbf{argmin}}\{\boldsymbol{L_S(h)}\}$ having small generalization error $L_{\mathcal{D},f}(h_S)$, with a condition that S is **big enough**. That mean the probability to select a bad sample at most equal δ

$$\boldsymbol{\mathcal{D}^m\left(\{S_x\colon L_{\mathcal{D},f}(h_S) > \varepsilon\}\right) = P_{S \sim \mathcal{D}^m}[x, L_{\mathcal{D},f}(h_S) > \varepsilon] \leq \delta : (PAC - Learning)}$$

1.1. ERM_H algorithm: Generalization bound of learning

- $P_{S\sim\mathcal{D}^m}[x, L_{\mathcal{D},f}(h_S) > \varepsilon] \leq \delta \Leftrightarrow P_{S\sim\mathcal{D}^m}[x, L_{\mathcal{D},f}(h_S) \leq \varepsilon] \geq 1-\delta$

- $[x, L_{\mathcal{D},f}(h_S) \leq \varepsilon] \cap [x, L_{\mathcal{D},f}(h_S) > \varepsilon] = \emptyset$

- $\rightarrow P([x, L_{\mathcal{D},f}(h_S) \leq \varepsilon] \cap [x, L_{\mathcal{D},f}(h_S) > \varepsilon]) = P(\emptyset) = 0$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Then we have :

- $P([x\, L_{\mathcal{D},f}(h_S) \leq \varepsilon] \cup [x, L_{\mathcal{D},f}(h_S) > \varepsilon])$= $P\,[x, L_{\mathcal{D},f}(h_S) \leq \varepsilon] + P\,[x, L_{\mathcal{D},f}(h_S) > \varepsilon] = 1$
- If $P_{S\sim\mathcal{D}^m}[x, L_{\mathcal{D},f}(h_S) > \varepsilon] \leq \delta$ then $P\,[x, L_{\mathcal{D},f}(h_S) \leq \varepsilon] = 1 - P\,[x, L_{\mathcal{D},f}(h_S) > \varepsilon] \geq 1-\delta$

1.1. ERM_H algorithm

Proof

Let

- $S = \{(x_i, y_i), i \in I\}$ and $\mathrm{S_x} = \{x_i, i \in I\}$, where $I = \{1, \dots, m\}$
- ε be a parameter that describes the prediction quality.

Lemmas:

- **Boole inequality** Let A and B be any events and $\mathcal{D}$ any distribution:
$$\mathcal{D}(A \cup B) = \mathcal{D}(A) + \mathcal{D}(B) - \mathcal{D}(A \cap B) \leq \mathcal{D}(A) + \mathcal{D}(B)$$
- **Independance** If A and B are two independent events, then:
$$\mathcal{D}(A \cap B) = \mathcal{D}(A).\mathcal{D}(B)$$

Note:

- $L_{\mathcal{D},f}(h_S) > \varepsilon \implies h_S$ is a bad hypothesis.
- $L_{\mathcal{D},f}(h_S) \leq \varepsilon \implies h_S$ is a good hypothesis.

Objective: Find an upper bound that limits the probability of having a bad sample (non representative):

$$\mathcal{D}^m\left(\left\{\mathrm{S_x}\colon L_{\mathcal{D},f}(h_S) > \varepsilon\right\}\right)$$

1.1. ERM_H algorithm

Consider:

- The set of bad samples enabling to select the bad hypothesis h_S:

$$S_B = \{\mathrm{S_x}\colon L_{\mathcal{D},f}(h_S) > \varepsilon\}$$

- The probability to select a bad sample:

$$\mathcal{D}^m(\{\mathrm{S_x}\colon L_{\mathcal{D},f}(h_S) > \varepsilon\})$$

- H_B the set of bad hypotheses:

$$H_B = \{h \in H \colon L_{\mathcal{D},f}(h) > \varepsilon\}$$

- M the set of misleading samples:

$$M = \{\mathrm{S_x}\colon \exists h \in H_B, L_S(h) = 0\}$$

Let's prove that:

$$\mathrm{S_B} \subseteq M$$

1.1. ERM_H algorithm

Let's fix any sample $S \in S_B$, this implies that $L_{\mathcal{D},f}(h_S) > \varepsilon$ and since the realizability assumption is respected, we will have $L_S(h_S) = 0$ with probability 1.

Then $\exists h \in H_B, and\ L_S(h) = 0$, This implies that $S \in M$

Hereby:

$$S_B = \{S_X \colon L_{\mathcal{D},f}(h_S) > \varepsilon\} \subseteq M$$

This implies that:

$$\mathcal{D}^m(S_B) = \mathcal{D}^m(\{S_X \colon L_{\mathcal{D},f}(h_S) > \varepsilon\}) \leq \mathcal{D}^m(M)$$

Notice that M can be represented by:

$$M = \bigcup_{h \in H_B} \{S_X \colon L_S(h) = 0\}$$

1.1. ERM_H algorithm

According to Boole Inequality, we obtain that :

$$\mathcal{D}^m\left(\left\{S_X : L_{\mathcal{D},f}(h_S) > \varepsilon\right\}\right) \leq \mathcal{D}^m\left(\bigcup_{h \in H_B} \{S_X \colon L_S(h) = 0\}\right) \leq \sum_{h \in H_B} \mathcal{D}^m(\{S_X \colon L_S(h) = 0\})$$

Let's fix a hypothesis $h \in H_B$ and limits by an upper bound the following expression:

$$\mathcal{D}^m(\{S_X \colon L_S(h) = 0\})$$

The event "$L_S(h) = 0$" is equivalent to $\forall i \in I, h(x_i) = f(x_i)$.

1.1. ERM_H algorithm

Hereby:

$$\mathcal{D}^m(\{S_X\colon L_S(h) = 0\}) = \mathcal{D}^m(\{S_X\colon \forall i \in I, h(x_i) = f(x_i)\})$$
$$= \mathcal{D}^m\Big(\big\{S_X\colon \big(h(x_1) = f(x_1)\big) \cap \cdots \cap \big(h(x_m) = f(x_m)\big)\big\}\Big)$$

Since x_i are sampled $(i.i.d.)$, so:

$$\mathcal{D}^m(\{S_X\colon L_S(h) = 0\}) = \prod_{i=1}^{m} \mathcal{D}(\{x_i\colon h(x_i) = f(x_i)\})$$

For each element in the training set, we have:

$$\mathcal{D}(\{x_i : h(x_i) = f(x_i)\}) = 1 - L_{\mathcal{D},f}(h) \leq 1 - \varepsilon$$

Because, $S_x \in M \Longrightarrow \exists h \in H_B \Longrightarrow L_{\mathcal{D},f}(h) > \varepsilon$.

So:

$$\prod_{i=1}^{m} \mathcal{D}(\{x_i\colon h(x_i) = f(x_i)\}) \leq (1 - \varepsilon)^m$$

Then:

$$\mathcal{D}^m(\{S_X\colon L_S(h) = 0\}) \leq (1 - \varepsilon)^m$$

1.1. ERM_H algorithm

We know that: $1-\varepsilon \leq e^{-\varepsilon}$.

So, for a fixed hypothesis $h \in H_B$:

$$\mathcal{D}^m(\{S_X\colon L_S(h)=0\}) \leq e^{-\varepsilon m}$$

For all $h \in H_B$, we have:

$$\sum_{h \in H_B} \mathcal{D}^m(\{S_X\colon L_S(h)=0\}) \leq \sum_{h \in H_B} e^{-\varepsilon m} = |H_B| e^{-\varepsilon}$$

So:

$$\mathcal{D}^m(S_B) = \mathcal{D}^m\left(\left\{S_X\colon L_{\mathcal{D},f}(h_S) > \epsilon\right\}\right) \leq \sum_{h \in H_B} \mathcal{D}^m(\{S_X\colon L_S(h)=0\}) \leq |H_B| e^{-\varepsilon m} \leq |H| e^{-\varepsilon m}$$

For S big enough $\exists \delta > 0$ small such that :

$$\boldsymbol{\mathcal{D}^m\left(\left\{S_X\colon L_{\mathcal{D},f}(h_S) > \epsilon\right\}\right) \leq |H| e^{-\varepsilon m} \leq \delta}$$

1.1. ERM_H algorithm

Each point of this circle presents a sample S of size m.

The set of bad samples S_{B_1} that generates the bad hypothesis h_1 such that $L_S(h_1) = 0$.

The set of bad samples S_{B_2} that generates the bad hypothesis h_2 such that $L_S(h_2) = 0$.

The set of bad samples S_{B_3} that generates the bad hypothesis h_3 such that $L_S(h_3) = 0$.

The set of bad samples S_{B_4} that generates the bad hypothesis h_4 such that $L_S(h_4) = 0$.

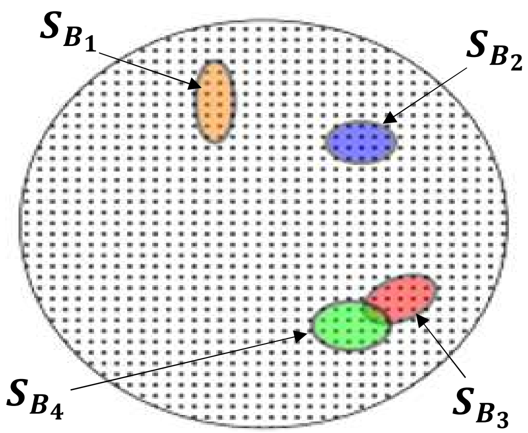
$$\vdots$$

Hence, we have the set of bad hypotheses:

$$H_B = \{h_1, h_2, h_3, h_4, \ldots\}$$

And the set of misleading samples:

$$M = \bigcup_{h \in H_B} \{S_X \colon L_S(h) = 0\} = \bigcup_{k \geq 1} S_{B_k}$$



1.1. ERM_H algorithm

We had demonstrated that:

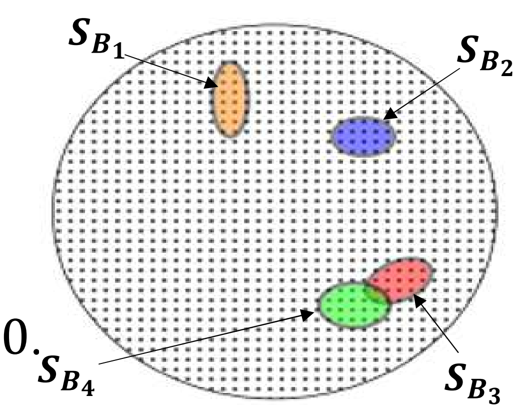
$$\mathcal{D}^m(S_B) \leq \mathcal{D}^m(M) \leq \sum_{k\geq 1} \mathcal{D}^m(S_{B_k}) \leq |H_B|e^{-\varepsilon m} \leq |H|e^{-\varepsilon m}$$

Such that $S_B \in \{S_{B_1}, S_{B_2}, S_{B_3}, S_{B_4}, \dots\}$

This means that the union of the colored ovals areas is at most equal to their sum. Which is bounded by $|H_B|$ times the maximum size of a colored oval.

If $m \to +\infty$ we have that:

$$\sum_{k\geq 1} \mathcal{D}^m(S_{B_k}) \to 0$$



It means that the size of the colored areas becomes small.

This implies that the probability of selecting a bad sample $(S \in S_B) \to 0$.

➢Therefore, we should collect the maximum number of training points.

1.1. ERM_H algorithm: S is **big enough**

Corollary:

let:

- ***H*** **is a finite hypotheses set,**
- $\delta \in [0,1]$ probability of selecting a bad sample and $\varepsilon > 0$ accuracy parameter
- m the size of S such that: $\boldsymbol{m \geq \frac{ln(|H|/\delta)}{\varepsilon}}$

So, $\forall(f, \mathcal{D})$ for which the **realizability assumption holds**.

With a probability at least equal to $(1-\delta)$ over the choice of an $(\boldsymbol{i.i.d.})$ sample S of size m, we have that for every hypothesis h_S selected by ERM_H it holds that :

$$\boldsymbol{P_{S \sim \mathcal{D}^m}(S_x, L_{\mathcal{D},f}(h_S) \leq \varepsilon) = \mathcal{D}^m\left(\left\{S_X\colon L_{\mathcal{D},f}(h_S) \leq \varepsilon\right\}\right) \geq 1-\delta}$$

$$\Longleftrightarrow$$

$$\boldsymbol{P_{S \sim \mathcal{D}^m}(S_x, L_{\mathcal{D},f}(h_S) > \varepsilon) = \mathcal{D}^m\left(\left\{S_X\colon L_{\mathcal{D},f}(h_S) > \varepsilon\right\}\right) \leq \delta}$$

1.1. ERM_H algorithm

Proof:

We have:

$$\mathcal{D}^m\left(\{S_X\colon L_{\mathcal{D},f}(h_S) > \epsilon\}\right) \leq |H|e^{-\varepsilon m}$$

We know that δ is a probability to select a bad sample.

We want this inequality to be less than δ :

$$|H|e^{-\varepsilon} \quad \leq \delta$$

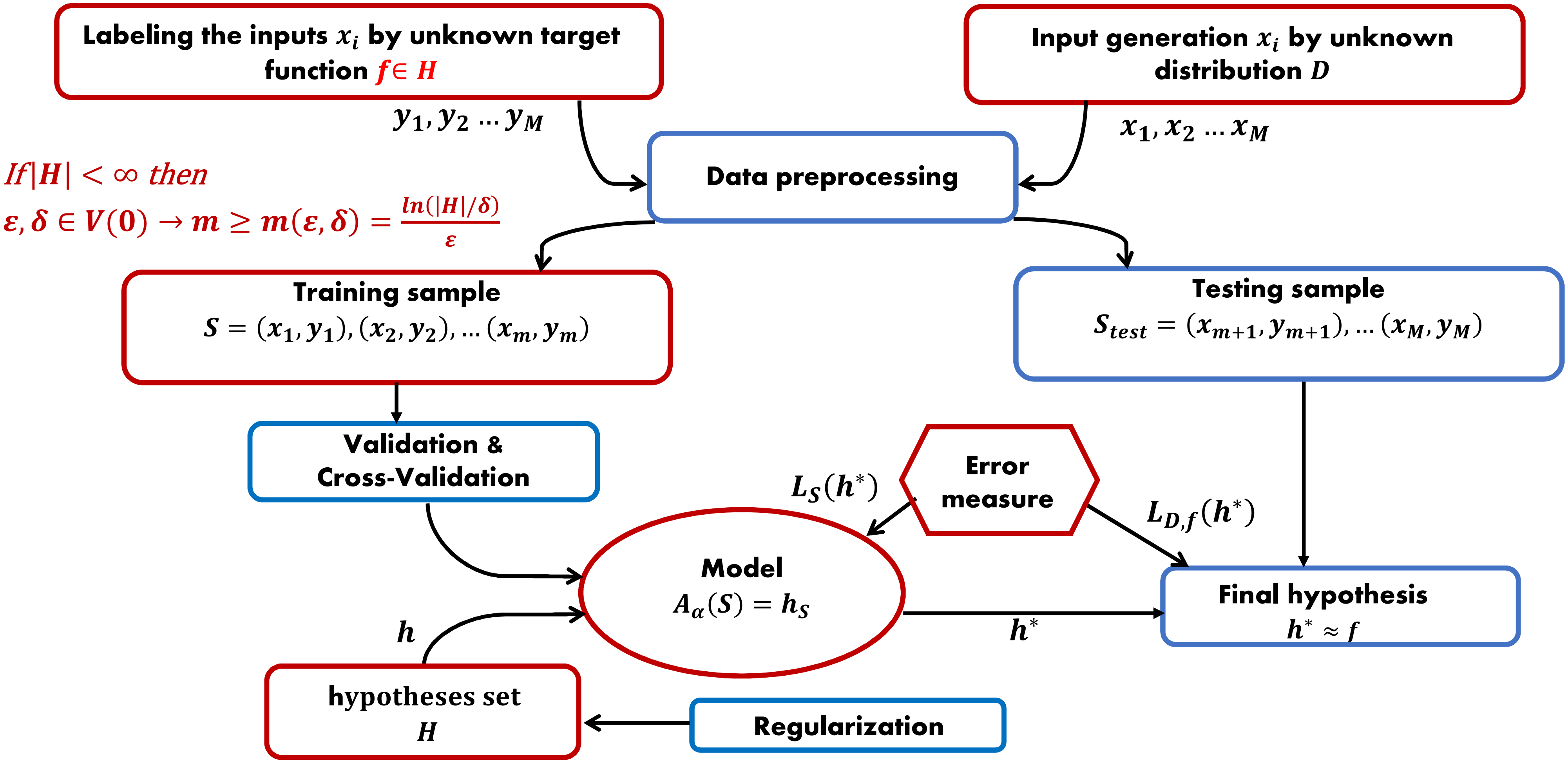
So:

$$m \geq \frac{ln(|H|/\delta)}{\varepsilon}$$

- If we have ε, m we can calculate δ
- If we have ε, δ we can calculate m

- The best of ε and δ when $\varepsilon, \delta \in V(0)$

Supervised Learning Passive Offline Algorithm (SLPOA)



1.2. PAC learning model

Definition:

H is **Probably Approximately Correct**, if there exist:

m_H: $(0,1)^2 \rightarrow \mathbb{N}$ and A_α having the following property:

- $\forall \varepsilon, \delta \in (0,1)$, $\forall\ (\mathcal{D}, f)$ and if the **realizability hypothesis is respected related** to H,$\mathcal{D}$ and $(f \in H)$.

So, if we run A_α(**learnable model**) on $m \geq m_H(\varepsilon, \delta)$(big enough) generated $(i.i.d.)$, such that S is selected by a probability at least equal to $(1-\delta)$, A_α will generate a hypothesis h_S such that:

$$L_{\mathcal{D},f}(h_S) \leq \varepsilon$$

In other words: $for\ all\ m \geq m_H(\varepsilon, \delta)$

$$P_{S\sim(\mathcal{D}^m,f)}\left[x\colon L_{\mathcal{D},f}(h_S) > \varepsilon\right] \leq \delta \Leftrightarrow P_{S\sim(\mathcal{D}^m,f)}\left[x\colon L_{\mathcal{D},f}(h_S) \leq \varepsilon\right] \geq 1-\delta.$$

$m_H(\varepsilon, \delta)$ minimum size of sample

1.2. PAC learning model

Definition:

The function m_H: $(0,1)^2 \rightarrow \mathbb{N}$ enables to determine the minimal number of data in S so that H follows a PAC learning, with accuracy ε and confidence δ.

The PAC definition owns two parameters :

- **Accuracy parameter** ε: It determines the distance between h and f. (**Approximately correct**).
- **Confidence parameter** δ: It determines the failure probability of the algorithm. (**Probably**).

Corollary: sample complexity $m_{\mathrm{H}}(\varepsilon, \delta)$

Any set of finite hypotheses H following PAC learning owns a sample complexity $m_{\mathrm{H}}(\varepsilon, \delta)$ such that:

$$m_{\mathrm{H}}(\varepsilon, \delta) = \left\lceil \frac{\ln\left(\frac{|\mathrm{H}|}{\delta}\right)}{\varepsilon} \right\rceil$$

1.3. SLPOA : General learning model

Generalization consists in relaxing two hypotheses:

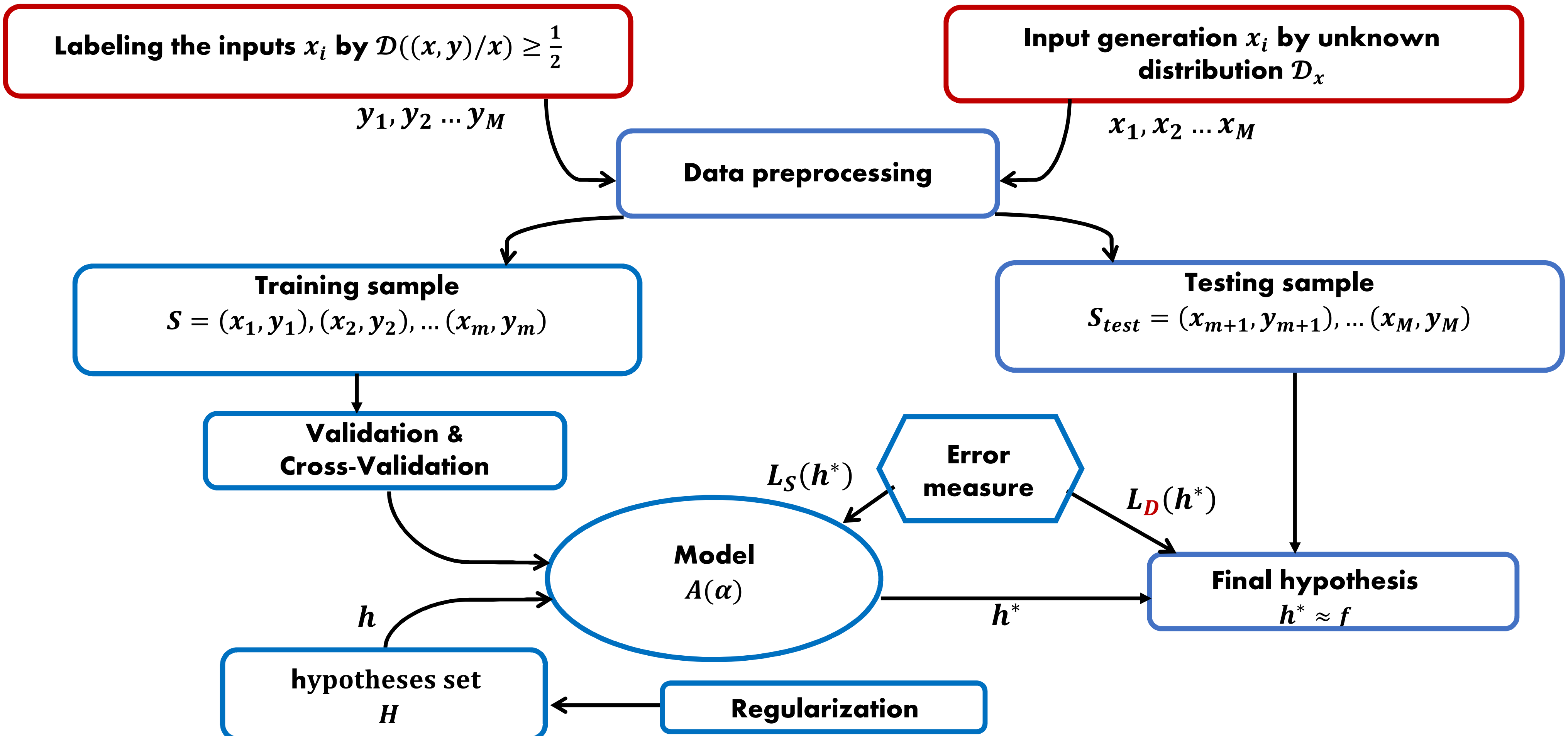
- The existence of f: we will consider that f is not deterministic then the **Realizability hypothesis isn't respected**
  - $\boldsymbol{\min_{h \in H} L_{\mathcal{D}}(h) \neq 0}$
  - $f \notin H$ with probability p ou $f \in H$ with probability 1-p

- The membership of f to H: we will remove the condition that $f \in H$.

So:

- D becomes a probability of joint distribution on $X \times Y$:
- $\mathcal{D}_x$: marginal probability distribution.
- $\mathcal{D}((x, y)/x)$: conditional probability distribution.

1.3. General learning model: SLPOA



1.3. General learning model

General error:

$$L_{\mathcal{D}}(h) \stackrel{\text{def}}{=} \mathbb{P}_{(x,y)\sim\mathcal{D}}[h(x) \neq y] \stackrel{\text{def}}{=} \mathcal{D}(\{(x,y): h(x) \neq y\})$$

Empirical error:

$$L_s(h) \stackrel{\text{def}}{=} \frac{|\{i \in I: h(x_i) \neq y_i\}|}{|S|}$$

Definition: Optimal hypothesis

Let the probability distribution $\mathcal{D}$ on $X \times \{0,1\}$, the best hypothesis $h^*: X \longrightarrow \{0,1\}$ is:

$$h^*(x) = \begin{cases} 1 \;\; if \;\; \mathcal{D}((x,1)|x) \geq \frac{1}{2} \\ 0 \;\; otherwise \;\; \mathcal{D}((x,1)|x) < \frac{1}{2} \end{cases} \Leftrightarrow h^*(x) = \begin{cases} 0 \;\; if \;\; \mathcal{D}((x,0)|x) \geq \frac{1}{2} \\ 1 \;\; otherwise \;\; \mathcal{D}((x,0)|x) < \frac{1}{2} \end{cases}$$

Notice:

This hypothesis is not accessible to the learning model A_α, because it requires the knowledge of $\mathcal{D}$.

1.3. General learning model

Definition: Agnostic PAC learning model

H follows agnostic PAC learning, if there exist $m_{\mathcal{H}}:\ (0,1)^2 \ \rightarrow \mathbb{N}$ and A_α having the following property:

$\forall \varepsilon, \delta \in (0,1), \forall\, \mathcal{D}$ on $X \times Y$, $\exists m_{\mathcal{H}}(\varepsilon, \delta)$ such that

We run A_α on $m \geq m_{\mathcal{H}}(\varepsilon, \delta)$ generated $(\boldsymbol{i.i.d.})$ such that S is selected with a probability at least $(1-\delta)$, A_α will generate the hypothesis h_S such that:

$$P_{S \sim \mathcal{D}^m}\left[\boldsymbol{L_D(h_S) \leq \min_{h \in H} L_{\mathcal{D}}(h) + \varepsilon}\right] \geq \mathbf{1} - \delta\ for\ all\ m \geq m_H(\varepsilon, \delta).$$

In other words:

$$P_{S \sim \mathcal{D}^m}\left[L_{\mathcal{D}}(h_S) > \boldsymbol{\min_{h \in H} L_{\mathcal{D}}(h)} + \varepsilon\right] \leq \delta\ \ for\ all\ m \geq m_H(\varepsilon, \delta)$$

1.3. General learning model

The general form of error:

Let l be a cost function, such that: $l: H \times Z \longrightarrow \mathbb{R}^{+}$ and $Z = X \times Y$

The general error of h: $\boldsymbol{L_{\mathcal{D}}(h) = \underset{z \sim \mathcal{D}}{\mathbf{E}}[l(h, z)]}$

The empirical error of h: $\boldsymbol{L_{S}(h) = \frac{1}{m} \sum_{i=1}^{m} l(h, z_i)}$

<table>
<tr><th>Classification</th><th>Regression</th></tr>
<tr><td>$$l(h,z) = \begin{cases} 1 \; si \; h(x) \neq y \\ 0 \; si \; h(x) = y \end{cases}$$
with:
$$z = (x,y) \in Z = X \times \{0,1\}$$
This function is also valid for the multinomial classification.</td><td>$$l(h,z) = (h(x) - y)^2$$
with:
$$z = (x,y) \in Z = X \times \mathbb{R}^{+}$$</td></tr>
</table>