Noisy targets

The 'target function' is not always a *function*

Consider the credit-card approval:

age	23 years
annual salary	$30,000
years in residence	1 year
years in job	1 year
current debt	$15,000
···	···

two 'identical' customers $\longrightarrow$ two different behaviors

Target 'distribution'

Instead of $y = f(\mathbf{x})$, we use target *distribution*:

$$P(y \mid \mathbf{x})$$

$(\mathbf{x}, y)$ is now generated by the joint distribution:

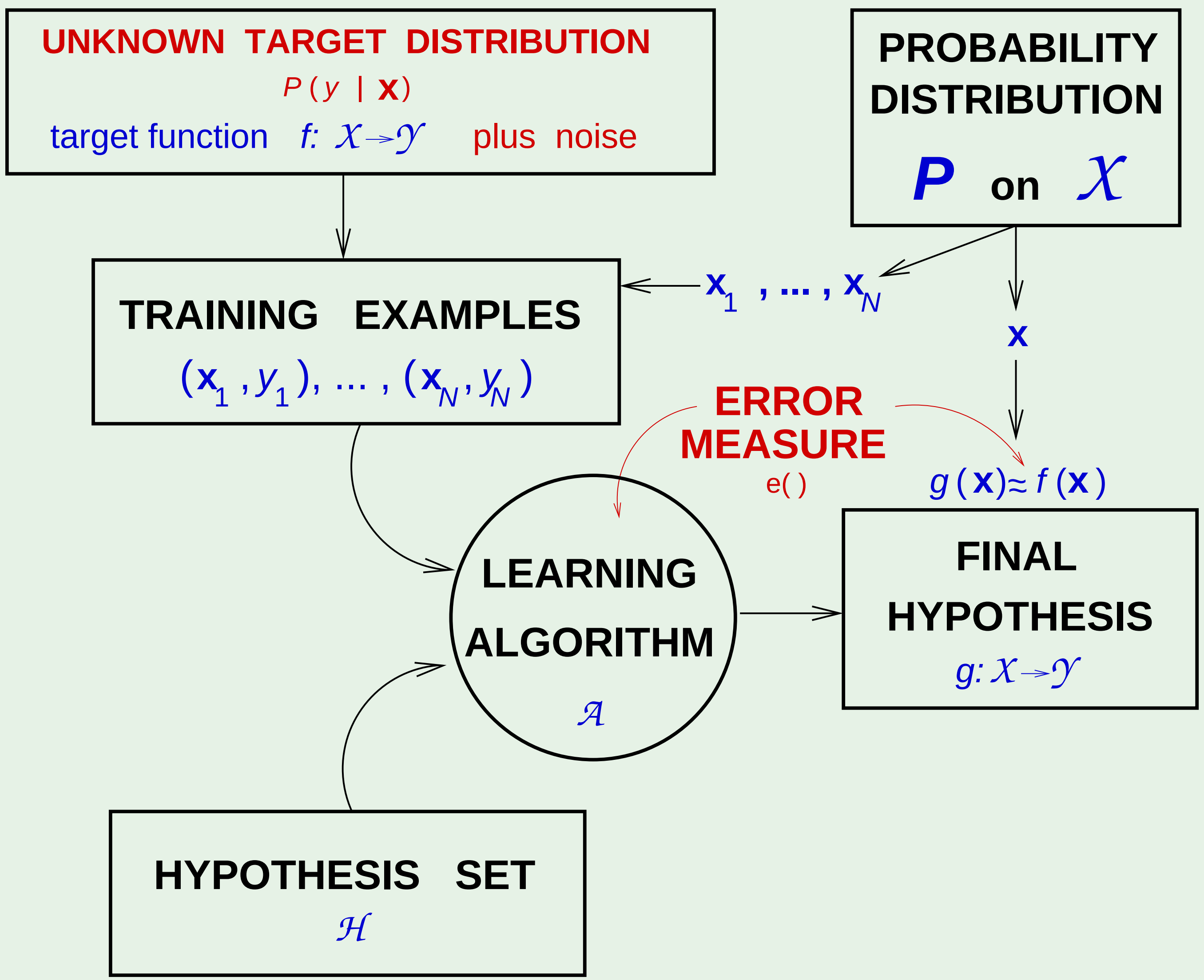
$$P(\mathbf{x})P(y \mid \mathbf{x})$$

Noisy target = deterministic target $f(\mathbf{x}) = \mathbb{E}(y|\mathbf{x})$ plus noise $y - f(\mathbf{x})$

Deterministic target is a special case of noisy target:

$$P(y \mid \mathbf{x}) \text{ is zero except for } y = f(\mathbf{x})$$

The learning diagram - including noisy target

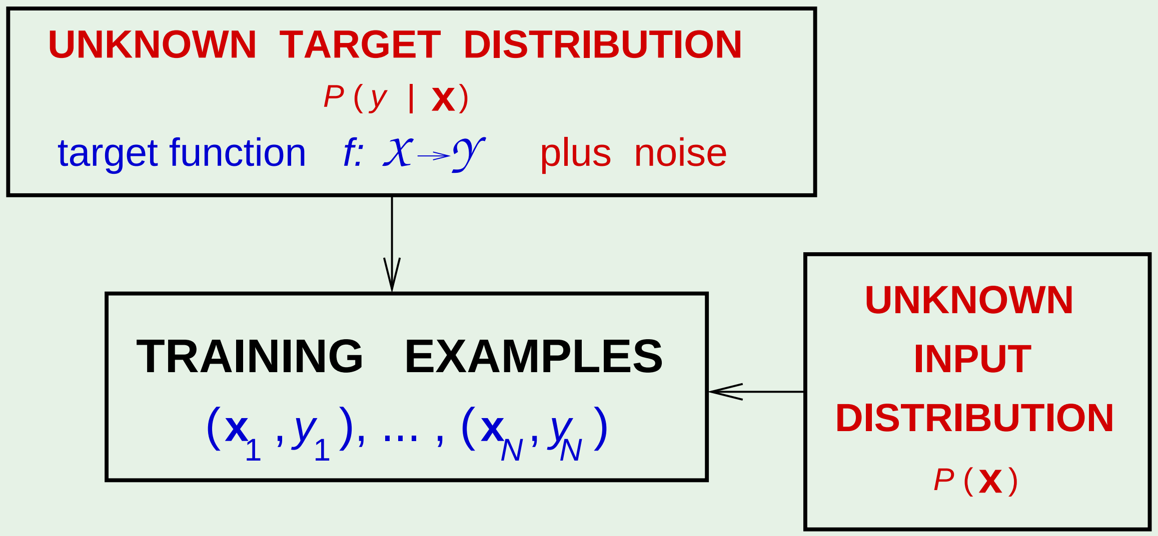


Distinction between $P(y|\mathbf{x})$ and $P(\mathbf{x})$

Both convey probabilistic aspects of $\mathbf{x}$ and y

The target distribution $P(y \mid \mathbf{x})$

is what we are trying to learn



The input distribution $P(\mathbf{x})$

quantifies relative importance of $\mathbf{x}$

Merging $P(\mathbf{x})P(y|\mathbf{x})$ as $P(\mathbf{x}, y)$

mixes the two concepts