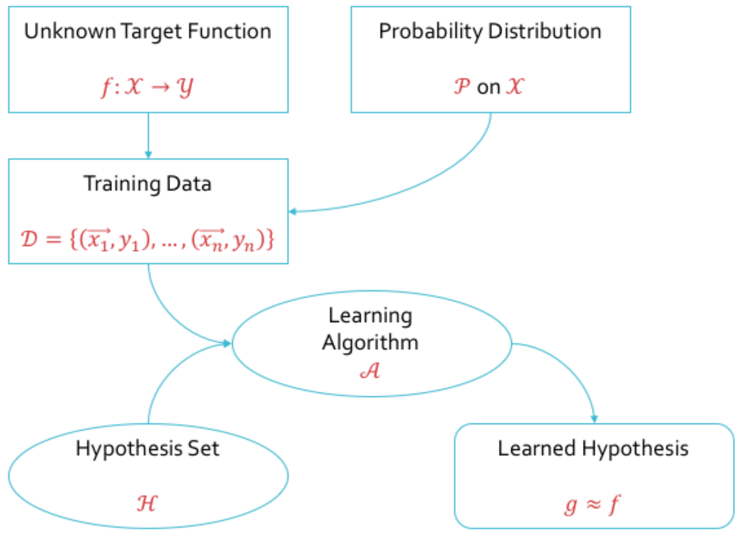
CSE 417T: Introduction to Machine Learning

Lecture 3: Error and Noise

Henry Chai

09/04/18

Recall


Unknown Target Function
$f: \mathcal{X} \rightarrow \mathcal{Y}$
Probability Distribution
$\mathcal{P}$ on $\mathcal{X}$
Training Data
$\mathcal{D} = \{(\overrightarrow{x_1}, y_1), \ldots, (\overrightarrow{x_n}, y_n)\}$
Learning Algorithm
$\mathcal{A}$
Hypothesis Set
$\mathcal{H}$
Learned Hypothesis
$g \approx f$

Error: is $h \approx f$?

- $E(h, f)$ is a function that measures how close h is to f: the smaller the better
- $g = \underset{h \in \mathcal{H}}{\operatorname{argmin}} E(h, f)$
- E is usually defined in terms of a pointwise error function $e(h, f, \vec{x})$
  - Binary error (classification): $e(h, f, \vec{x}) = [\![f(\vec{x}) \neq h(\vec{x})]\!]$
  - Squared error (regression): $e(h, f, \vec{x}) = \left(f(\vec{x}) - h(\vec{x})\right)^2$

3

Error: is $h \approx f$?

- In-sample error: $E_{in}(h) = \frac{1}{n}\sum_{i=1}^{n} e(h, f, \vec{x_i})$
- Out-of-sample error: $E_{out}(h) = \mathbb{E}_{x\sim\mathcal{P}}[e(h, f, \vec{x})]$

		$f(\vec{x})$	
		+1	-1
$h(\vec{x})$	+1	No error	False positive
	-1	False negative	No error

Error: Classification

5

There are two types of classification errors: false positives a.k.a. type 1 errors and false negatives a.k.a. type 2

Error: Classification

- Fingerprint recognition:
  - Inputs are fingerprints
  - Outputs: +1 means "you", -1 means "not you"

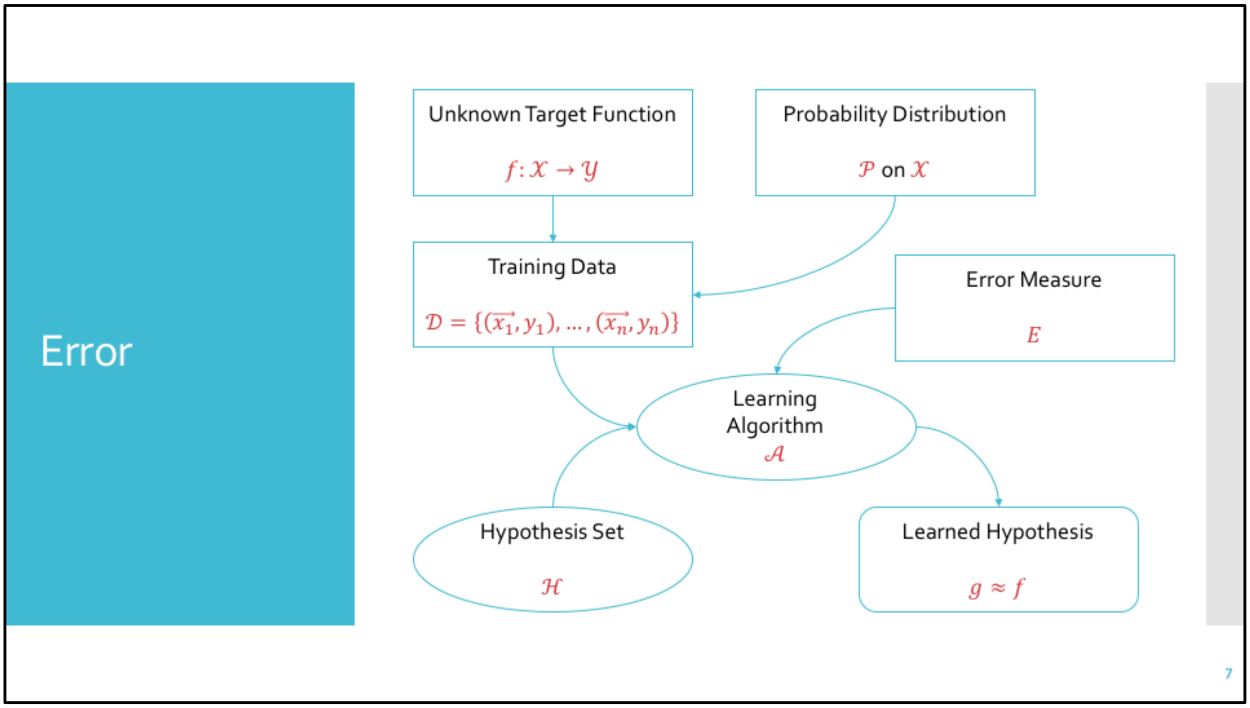
- For personalized coupons:

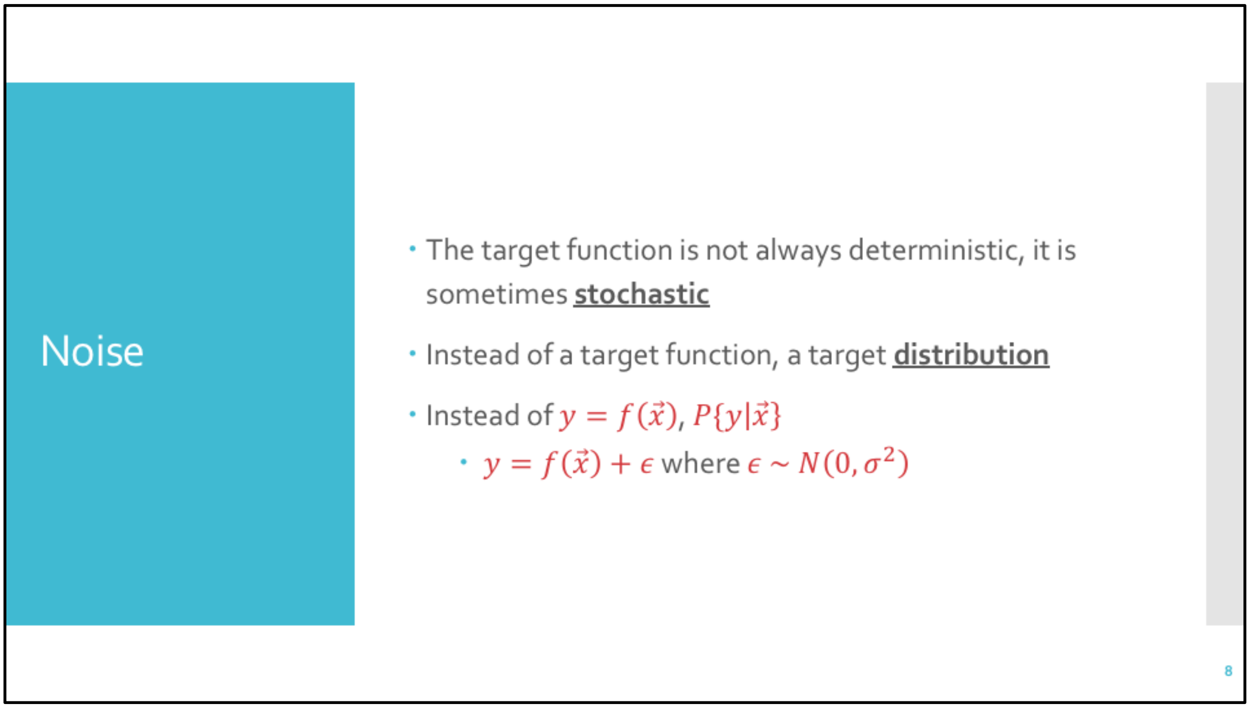
		$f(\vec{x})$	
		+1	-1
$h(\vec{x})$	+1	0	1
	-1	100	0

For unlocking phones:

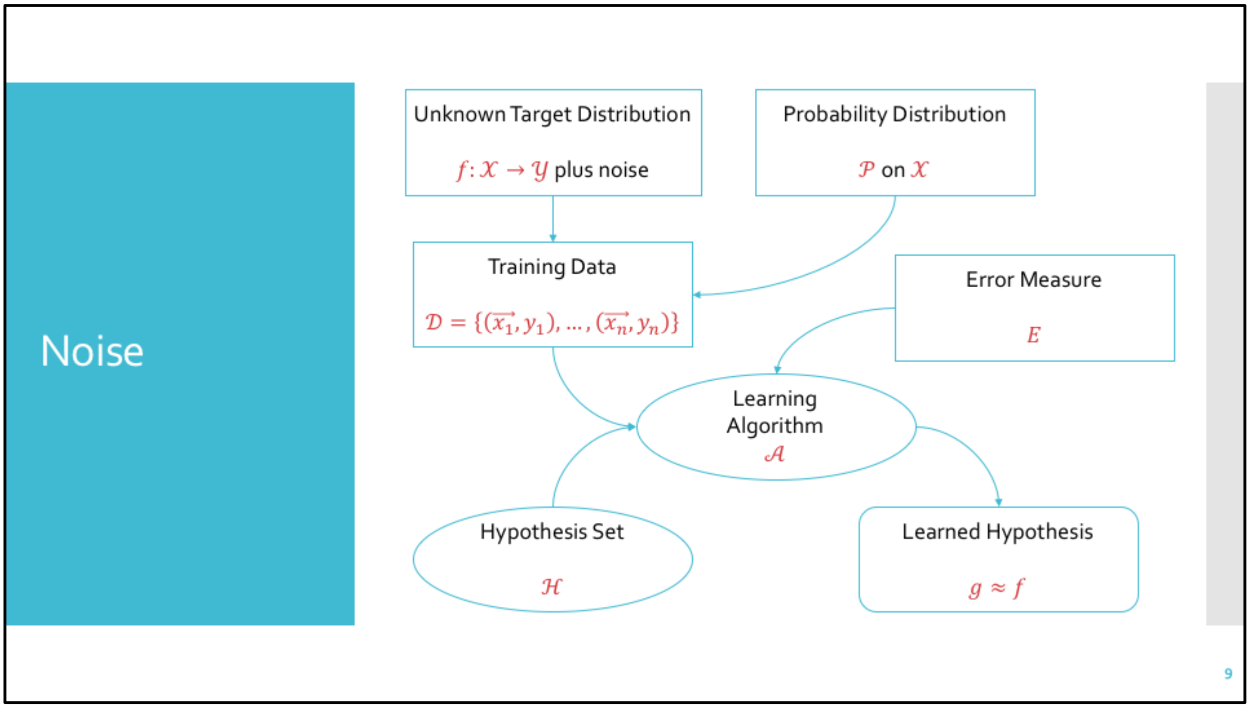
		$f(\vec{x})$	
		+1	-1
$h(\vec{x})$	+1	0	1000
	-1	1	0

The error function is user-specified and the optimal error function will depend on the application. If you don't have any particular reason to use one specific error measure over another, pick something computational-friendly i.e. squared error for regression


Error
Unknown Target Function
$f: \mathcal{X} \rightarrow \mathcal{Y}$
Probability Distribution
$\mathcal{P}$ on $\mathcal{X}$
Training Data
$\mathcal{D} = \{(\vec{x_1}, y_1), \ldots, (\vec{x_n}, y_n)\}$
Error Measure
E
Learning Algorithm
$\mathcal{A}$
Hypothesis Set
$\mathcal{H}$
Learned Hypothesis
$g \approx f$
7



Going back to our machine learning example from the first lecture: consider two people who came into the bank with the exact same attributes (as far as the bank can see) but had different results (one good and one bad)



Note that adding noise to the formal setup does not cause a loss of generality; you can still describe noise-less or deterministic target functions using this setup via Dirac delta functions