ECE595 / STAT598: Machine Learning I
Lecture 22.2: Is Learning Feasible?

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Outline

Today's Lecture:

- What constitutes a learning problem?
  - Training and testing samples
  - Target and Hypothesis function
  - Learning Model
- Is learning feasible?
  - An example
  - The power of probability
- Training versus Testing
  - In-sample error
  - Out-sample error
  - Probability bound

Is Learning Feasible?

In-sample and **Out-sample**:

- In-sample: Training Data
- Out-sample: Testing Data

When can we claim "learning is feasible"?

Suppose we have a training set $\mathcal{D}$, can we learn the target function f?

- "Learn" means: I use the data you give me to come up with an f
- "Successful" means: All in-samples are correctly predicted
- And all out-samples are also correctly predicted
- If YES, then we are in business.
- Learning is feasible!
- If NO, then we can go home and sleep.
- There is just no way to learn f from $\mathcal{D}$.

Example

- Let $\mathcal{X} = \{0, 1\}^3$
- Each $\boldsymbol{x} \in \mathcal{X}$ is a binary vector
- E.g., $\boldsymbol{x} = [0,\ 0,\ 1]^T$ or $\boldsymbol{x} = [1,\ 0,\ 1]^T$
- How many possible vectors are there? $2^3 = 8$
- Call them $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_8$
- There is a target function f
- f maps every $\boldsymbol{x}$ to a y
- $y \in \{+1, -1\}$
- E.g., $f([0, 0, 1]) = +1$, $f([0, 1, 1]) = -1$, etc.
- How many possible f's?
- You can think of f as a 8-bit vector
- E.g., $f = [+1, -1, -1, -1, +1, +1, +1, -1]$.
- So there are $2^8 = 256$ possible f's.

Example

- We have 8 input vectors: $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_8\}$
- We have 256 hypotheses: $\mathcal{H} = \{h_1, \ldots, h_{256}\}$
- Is learning feasible?
- Give me a subset $\mathcal{D} \subset \mathcal{X}$, can I find a hypothesis $g \in \mathcal{H}$ such that $g = f$?
- Suppose here is what you are given: $\circ = -1$, $\bullet = +1$. You know 6 out of 8. These are the training data.

$\boldsymbol{x}_n$			y_n
0	0	0	$\circ$
0	0	1	$\bullet$
0	1	0	$\bullet$
0	1	1	$\circ$
1	0	0	$\bullet$
1	0	1	$\circ$
1	1	0	?
1	1	1	?

Possibility 1

$\mathbf{x}_n$			y_n
0	0	0	∘
0	0	1	•
0	1	0	•
0	1	1	∘
1	0	0	•
1	0	1	∘
1	1	0	∘
1	1	1	∘

- One 1's will give me •; Others give me ∘
- So the last two entries should be ∘

Possibility 2

$\mathbf{x}_n$			y_n
0	0	0	$\circ$
0	0	1	$\bullet$
0	1	0	$\bullet$
0	1	1	$\circ$
1	0	0	$\bullet$
1	0	1	$\circ$
1	1	0	$\circ$
1	1	1	$\bullet$

- Odd numbers of 1's give me $\bullet$
- Even numbers of 1's give me $\circ$
- So [1 1 0] should be $\circ$
- So [1 1 1] should be $\bullet$

All the Possibilities

	$\mathbf{x}_n$		y_n	g	f_1	f_2	f_3	f_4
0	0	0	○	○	○	○	○	○
0	0	1	●	●	●	●	●	●
0	1	0	●	●	●	●	●	●
0	1	1	○	○	○	○	○	○
1	0	0	●	●	●	●	●	●
1	0	1	○	○	○	○	○	○
1	1	0		○/●	○	●	○	●
1	1	1		○/●	○	○	●	●

- f_1, f_2, f_3, f_4 are the only hypotheses you need to consider
- You just don't know which one out of the four to choose!
- You won't do better the random guess.
- So you haven't learned anything from the training data.
- Learning is infeasible.

The Power of Probability