Machine Learning from Data

Lecture 5: Spring 2021

Today's Lecture

- Training Vs Testing
  - The Two Questions of Learning
  - Theory of Generalization (Ein ≈ Eout)
  - An Effective Number of Hypotheses

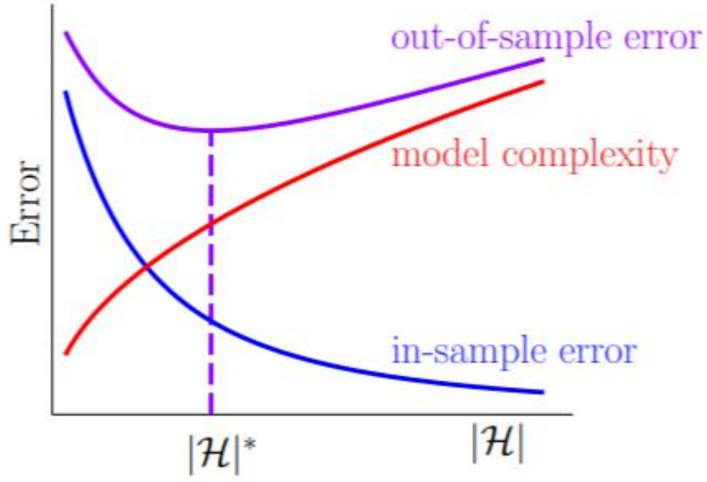
1. Can we make sure that $E_{\text{out}}(g)$ is close enough to $E_{\text{in}}(g)$?
2. Can we make $E_{\text{in}}(g)$ small enough?

The Hoeffding *generalization bound*:

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \underbrace{\sqrt{\frac{1}{2N} \ln \frac{2|\mathcal{H}|}{\delta}}}_{\text{generalization error bar}}$$

E_{in}: training (eg. the practice exam)

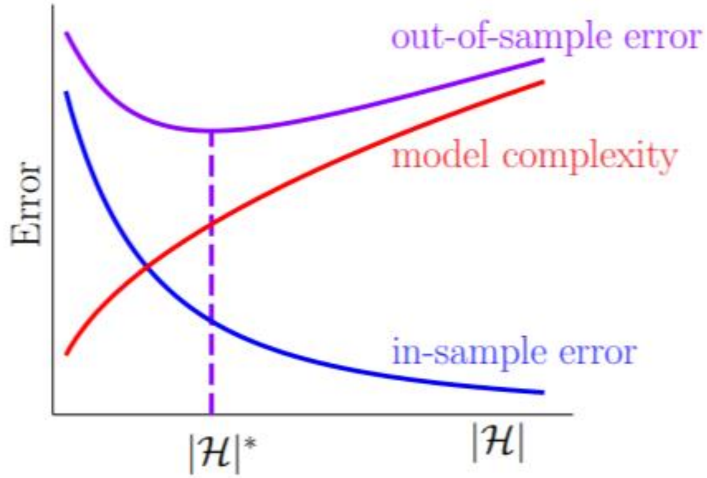
E_{out}: testing (eg. the real exam)



There is a tradeoff when picking $|\mathcal{H}|$.

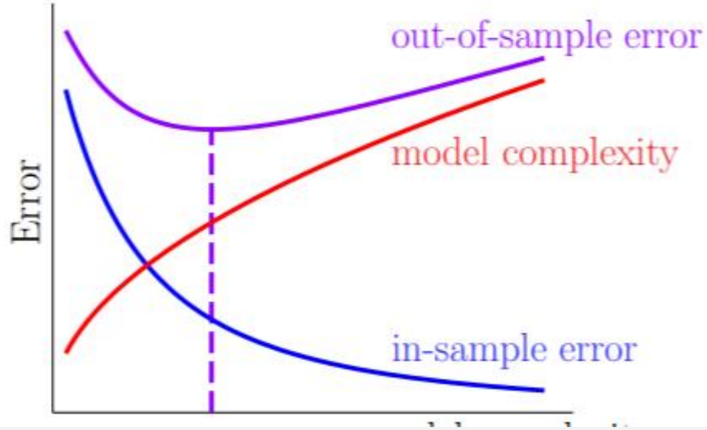
The Two Questions of Learning

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{1}{2N} \ln \frac{2|\mathcal{H}|}{\delta}}$$

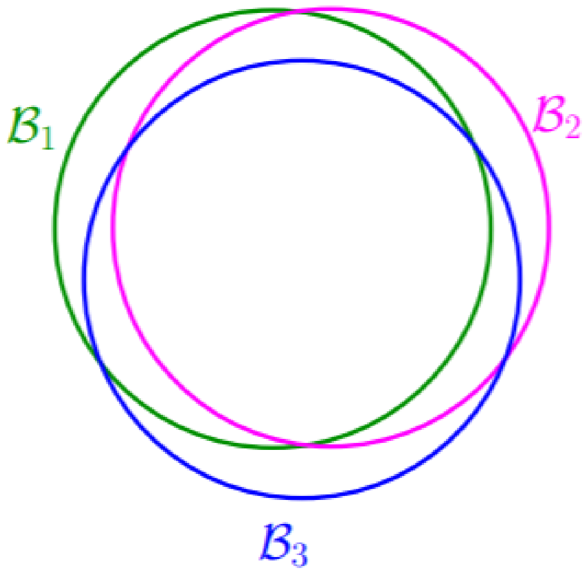


$$\downarrow$$

$$E_{\text{out}}(g) \leq E_{\text{in}}(g) + \sqrt{\frac{8}{N} \ln \frac{4m_{\mathcal{H}}}{\delta}}$$



Theory of Generalization



- $\mathcal{B}_m$ are events (sets of outcomes); they can overlap.
- If the $\mathcal{B}_m$ overlap, the union bound is loose.
- If many h_m are similar, the $\mathcal{B}_m$ overlap.
- There are "effectively" fewer than $|\mathcal{H}|$ hypotheses,.
- We can replace $|\mathcal{H}|$ by something smaller.

Meaning of cardinality of H?

How did H come in?

Effective Number of Hypothesis

- We need a way to measure the diversity of H.
- A simple idea:
  - Fix any set of N data points.
  - If H is diverse it should be able to implement all functions . . . on these N points.

The Growth Function

•

Another Example

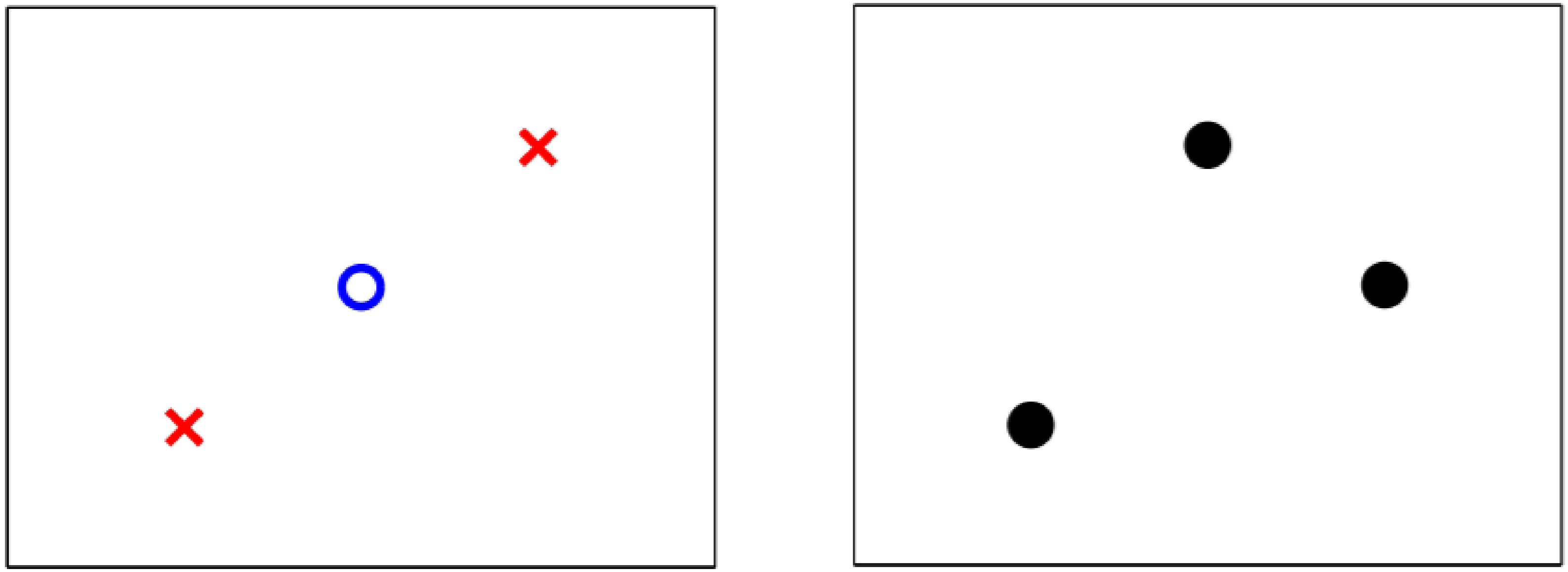
•

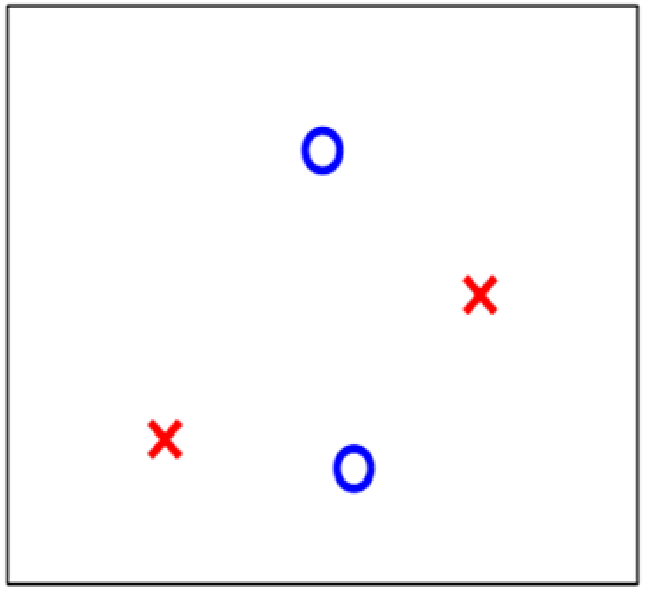
Dichotomies in 2-D Perceptron

•

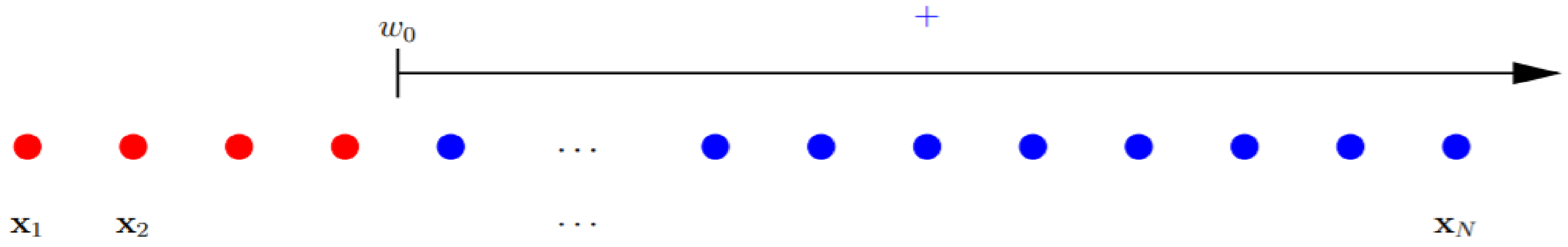
Example - Growth Functions

2-D Perceptron Model





1-D Positive Ray



Positive Rectangles in 2-D

-

Summarize

	N 1	2	3	4	5	$\cdots$
2-D perceptron	2	4	8	14	$\cdots$	
1-D pos. ray	2	3	4	5	$\cdots$	
2-D pos. rectangles	2	4	8	16	$< 2^5$	$\cdots$

- $m_{\mathcal{H}}(N)$ drops below 2^N – there is hope for the generalization bound.
- A break point is any n for which $m_{\mathcal{H}}(n) < 2^n$.

Definition: Shatter a Data Set

-

Combinatorial Puzzle

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
○	●	●

2 points are shattered

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
○	●	●

No Pair of Points is Shattered

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

If N = 4 how many possible dichotomys with no 2 points shattered?

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$
○	○	○
○	○	●
○	●	○
●	○	○

$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$
○	○	○	○
○	○	○	●
	⋮		

Thanks!