Experimental Design for Machine Learning on Multimedia Data
Lecture 6

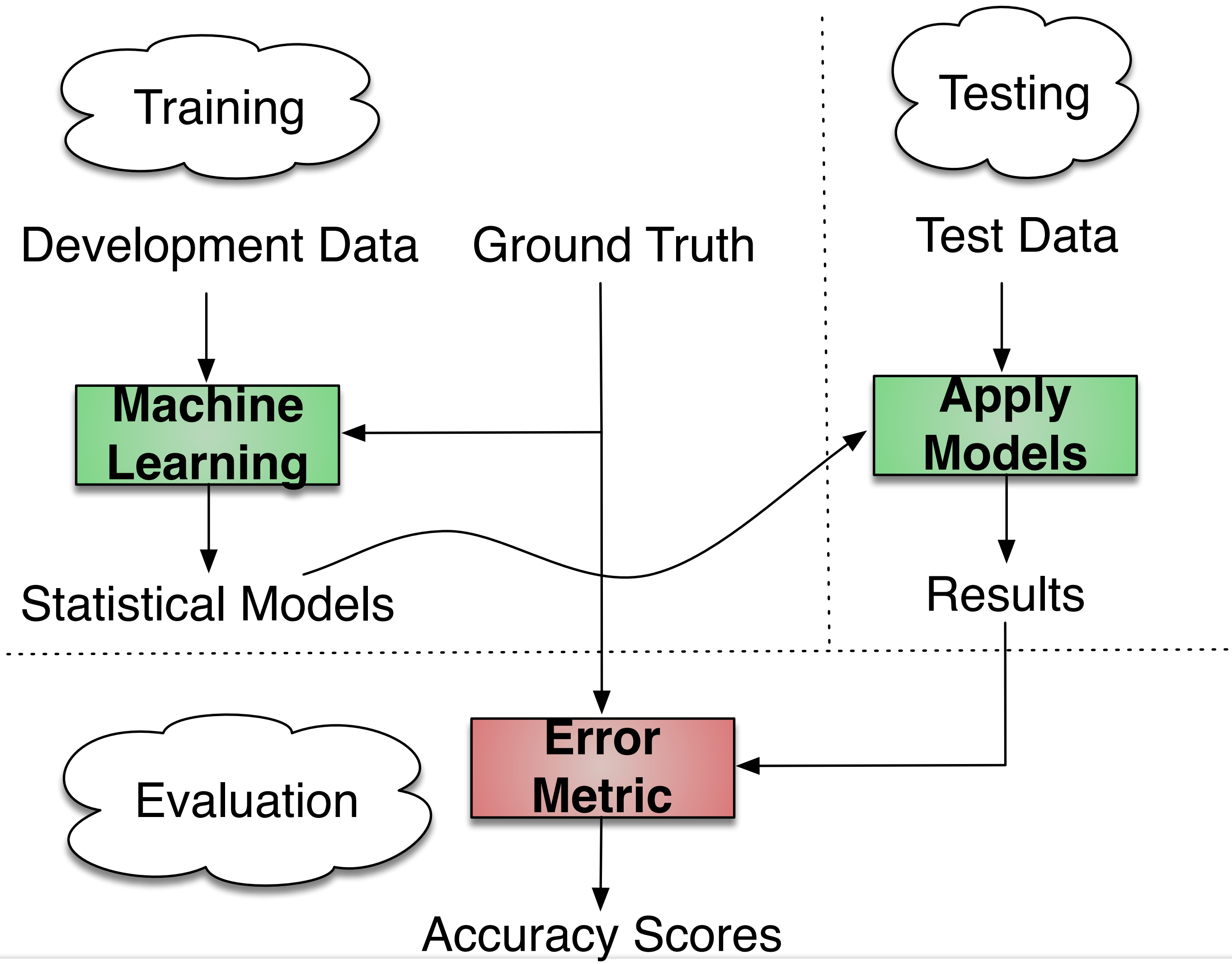
Dr. Gerald Friedland,
fractor@eecs.berkeley.edu

Website: http://www.icsi.berkeley.edu/~fractor/spring2019/

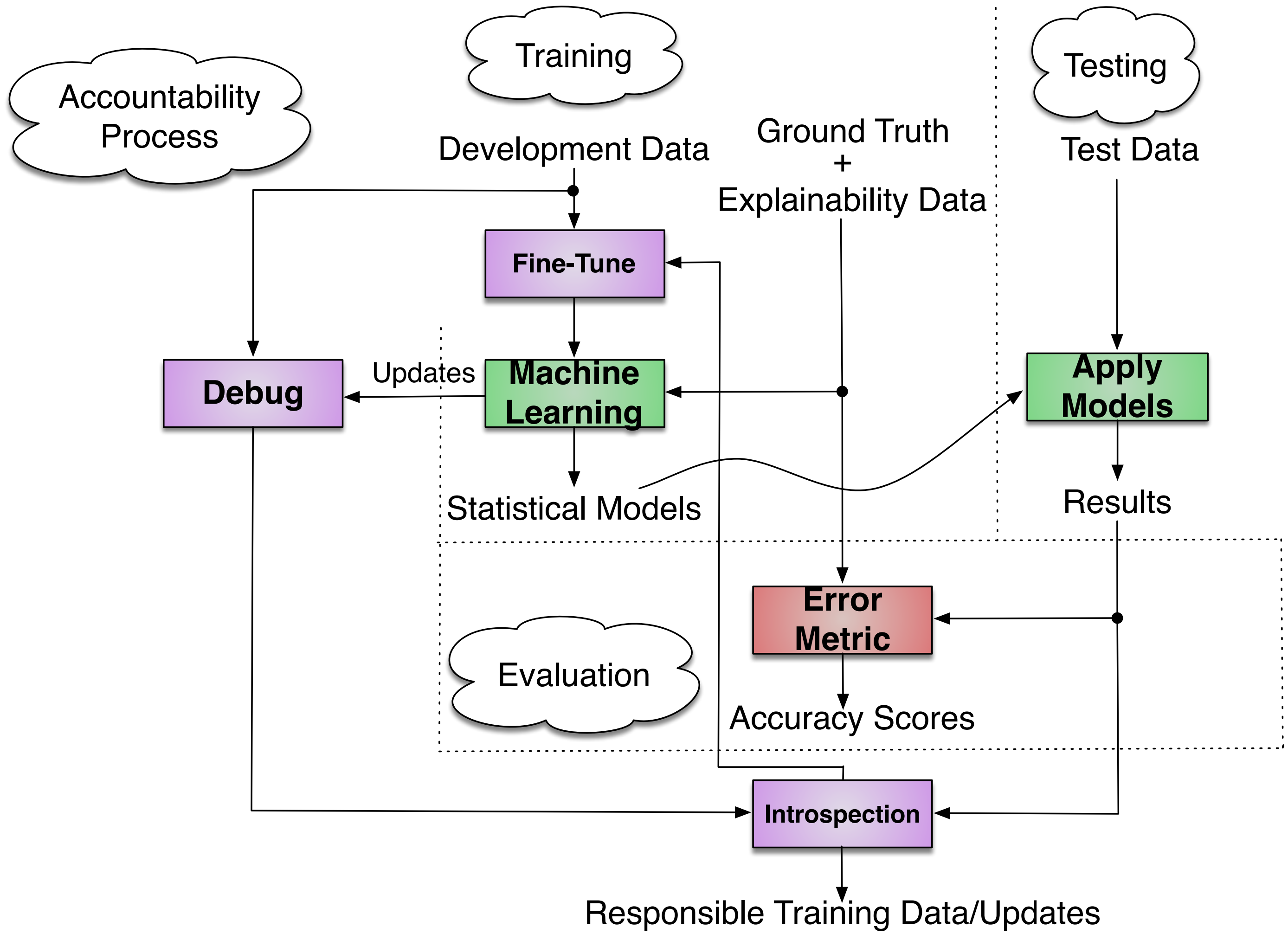
Projects

- Thank you for your submissions
- Feedback individually…

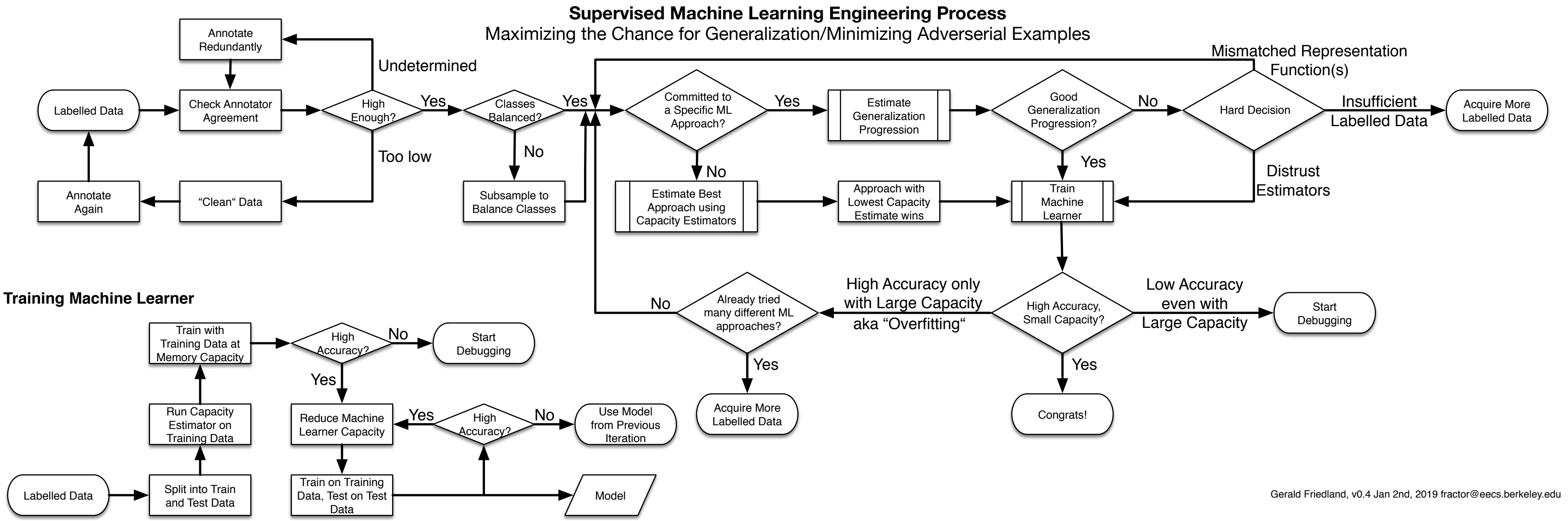
Recap: Generic Project Workflow for Accuracy



Generic Project Workflow for Explainability (DARPA)



Generic Project Workflow for Generalization



Generalization

1) What is generalization (intuitively)?

2) Formal definition

3) Why is it important

Abstraction

- **Detail removal**

  "The act of leaving out of consideration one or more properties of a complex object so as to attend to others."

- **Generalization**

  "The process of formulating general concepts by abstracting common properties of instances"

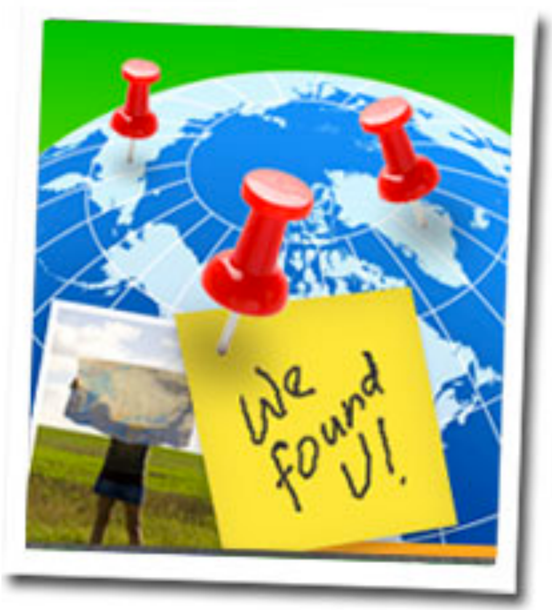
- **Technical terms:**
  Compression, Quantization, Clustering, Unsupervized Learning

Henri Matisse "Naked Blue IV"

Experiment

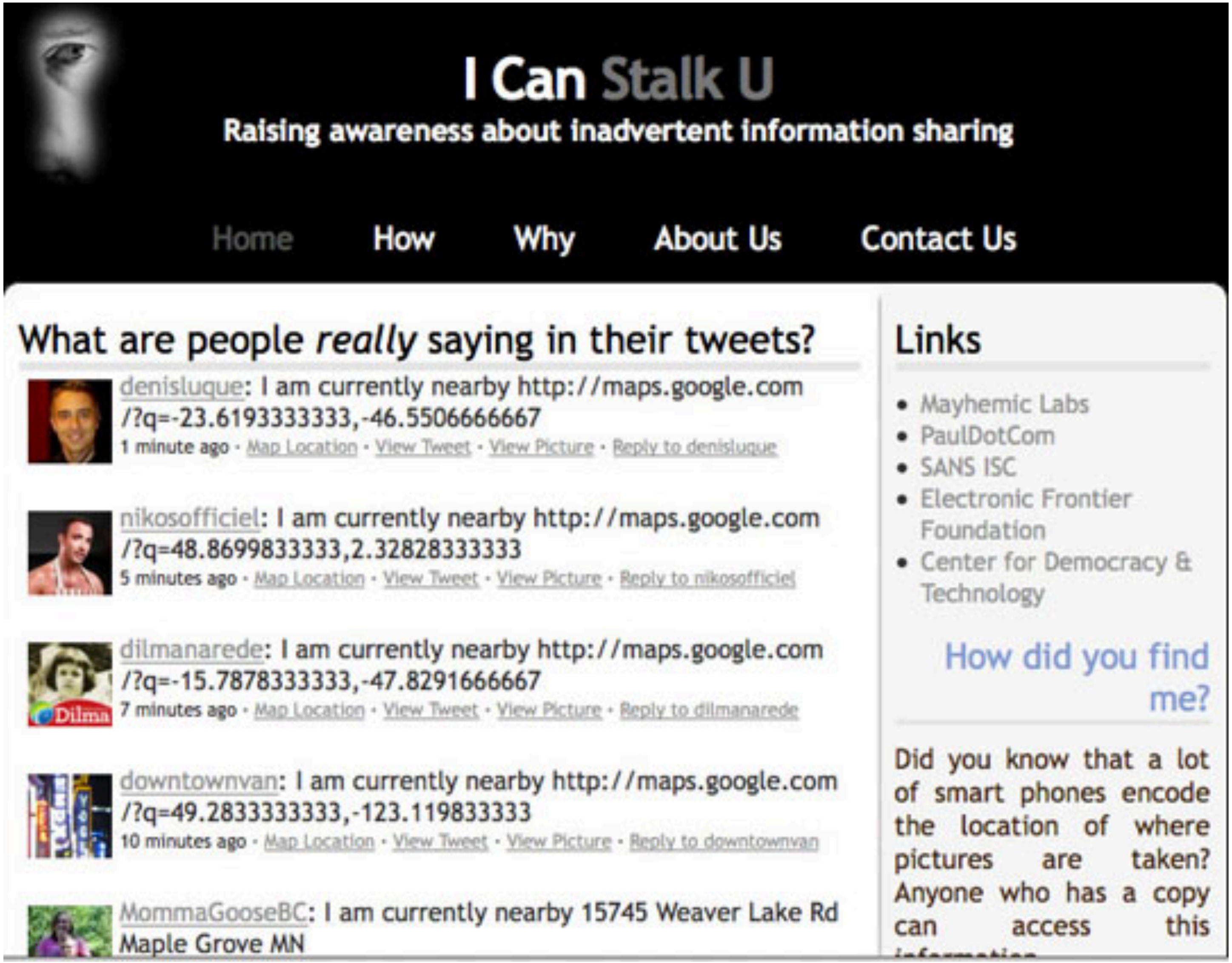
Where are you from?

Possible Answers:
- China
- California
- The Bay Area
- San Mateo
- 1947 Center Street, Berkeley, CA
- 37.8693° N, 122.2696° W



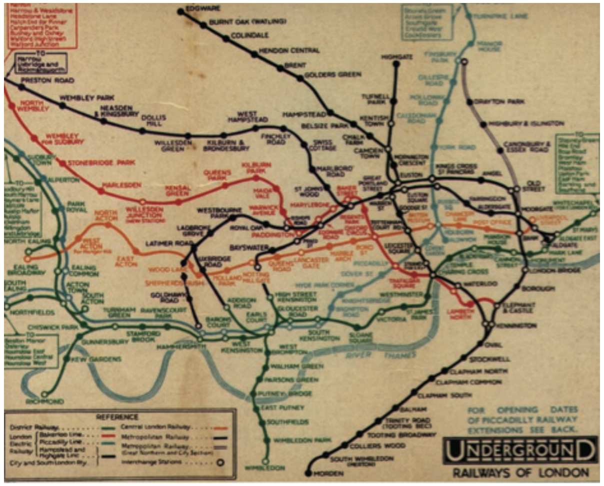
All correct but different levels of abstraction!

Abstraction gone wrong!



Detail Removal (in Data Science)

- You'll want to look at only the interesting data, leave out the details, zoom in/out…
- Abstraction is the idea that you focus on the essence, the cleanest way to map the messy real world to one you can build
- Experts are often brought in to know what to remove and what to keep!



The London Underground 1928 Map & the 1933 map by Harry Beck.

The Power of Abstraction, Everywhere!

- **Examples:**
  - **Functions (e.g., sin x)**
  - **Hiring contractors**
  - **Application Programming Interfaces (APIs)**
  - **Technology (e.g., cars)**

- **Amazing things are built when these layer**
  - **And the abstraction layers are getting deeper by the day!**

We only need to worry about the interface, or specification, or contract NOT how (or by whom) it's built

Above the abstraction line

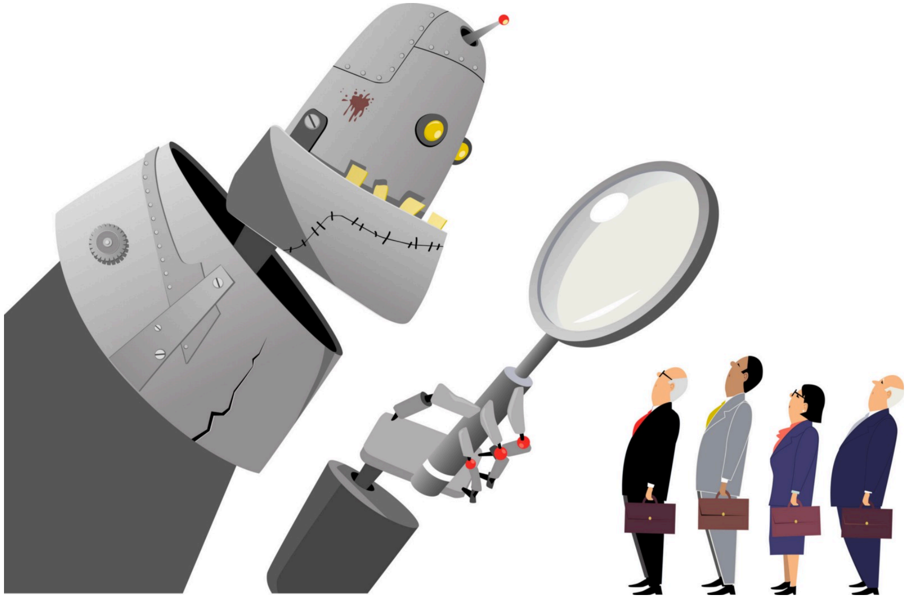
Abstraction Barrier (Interface)
(the interface, or specification, or contract)

Below the abstraction line

This is where / how / when / by whom it is actually built, which is done according to the interface, specification, or contract.

Abstraction: Pitfalls

- Abstraction is not universal without loss of information (mathematically provable). This means, in the end, the complexity can only be "moved around"
- Abstraction makes us forget how things actually work and can therefore hide bias. Example: AI and hiring decisions.



- Abstraction makes things special and that creates dependencies. Dependencies grow longer and longer over time and can become unmanageable.

Formal Definition: Generalization

Definition 0 (generalization in ML):

$$\forall x, x' \exists \delta \text{ such that } |x - x'| < \delta \implies f(x) = f(x')$$

$x \in$ Training data

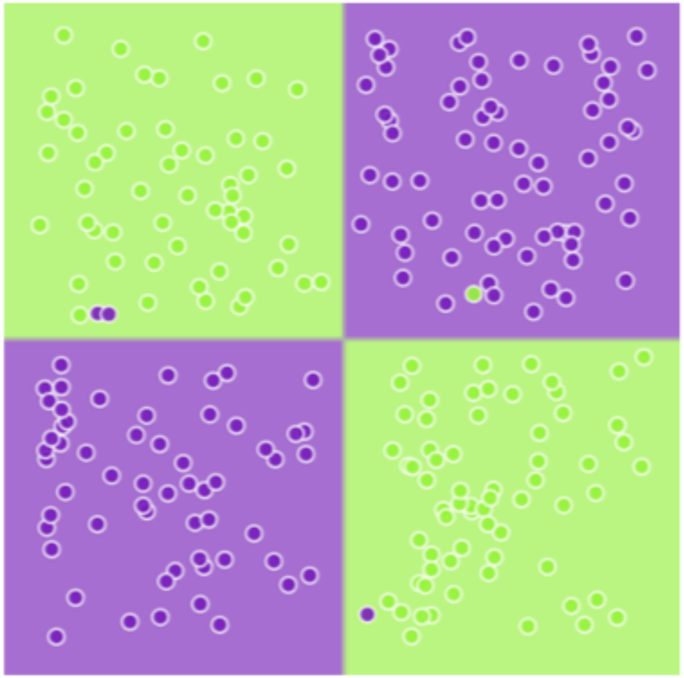
$x' \in$ Test data

$| \circ |$ semi-metric

f machine learner

Adversarial Example

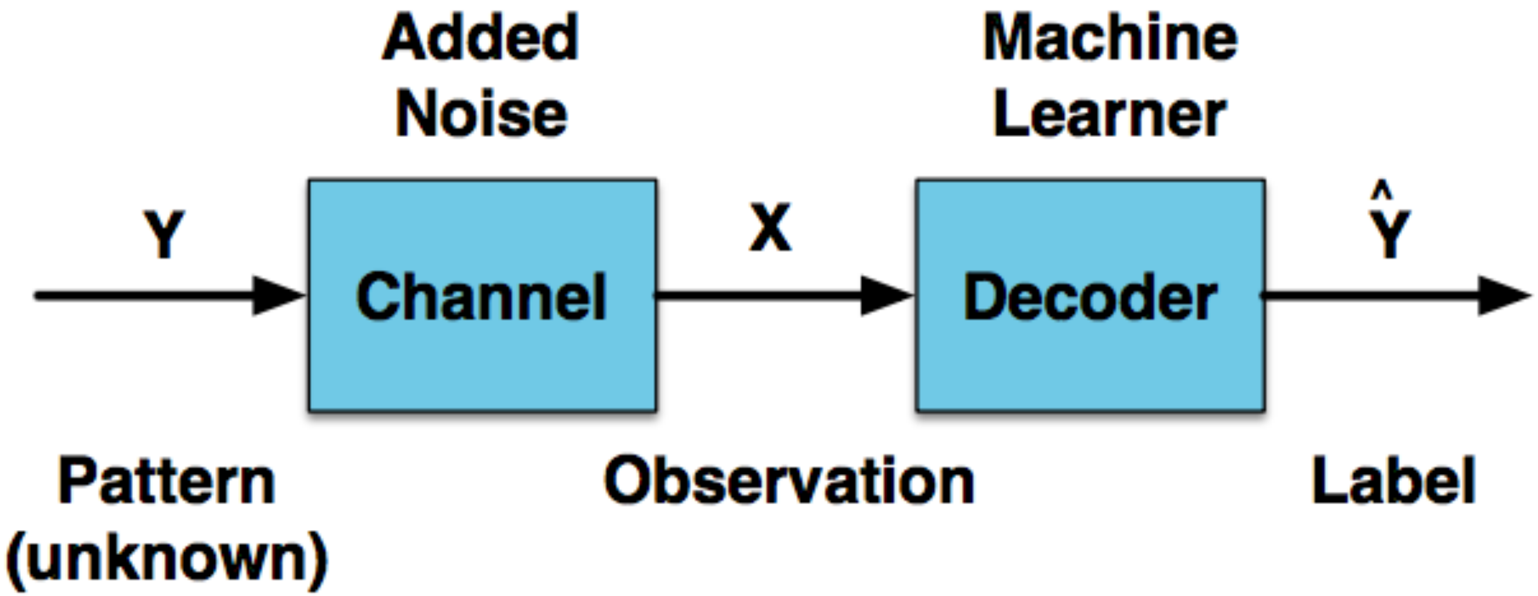
"XOR plus noise"



Definition 1 (Adversarial example (Wang et al., 2016)). *Given an ML model $f(\cdot)$ and a small perturbation δ, we call x' an adversarial example if there exists x, an example drawn from the benign data distribution, such that $f(x) \neq f(x')$ and $\|x - x'\| \leq \delta$.*

A Thermodynamic/Information Model for deterministic ML

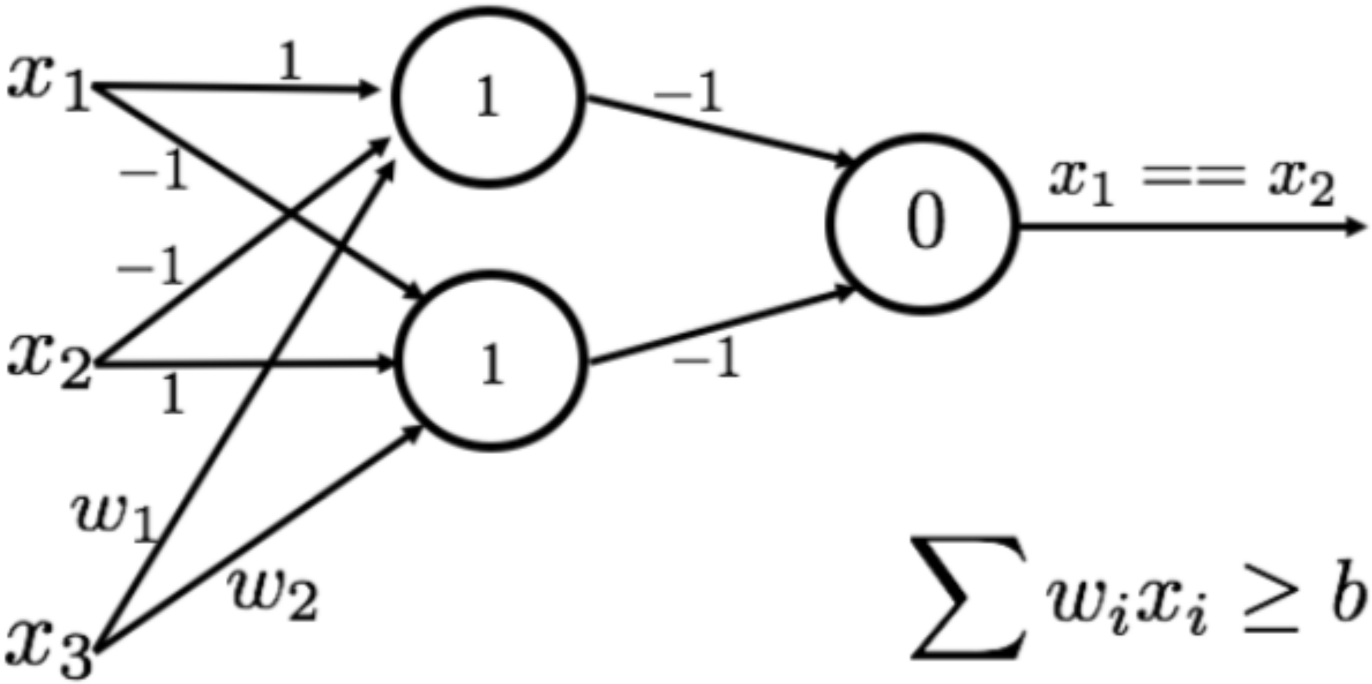
- **Machine Learning resets bits introduced by noise.**



- **Machine Learning denoises an unknown pattern.**

Adversarial Examples are caused by Redundancies: Logic View

Boolean Equality Network with a redundant variable.



Adversarial Examples are caused by Redundancies: Logic View

Exhaustive experimentation on NXOR network:

#Adv. Examples	#Allowing Edges	Potential
0	0	0
4	1	1
4	1	1
4	2	0
4	1	1
4	1	1
6	2	0
6	2	2
6	2	2

Adversarial Examples are caused by Redundancies: Experiments

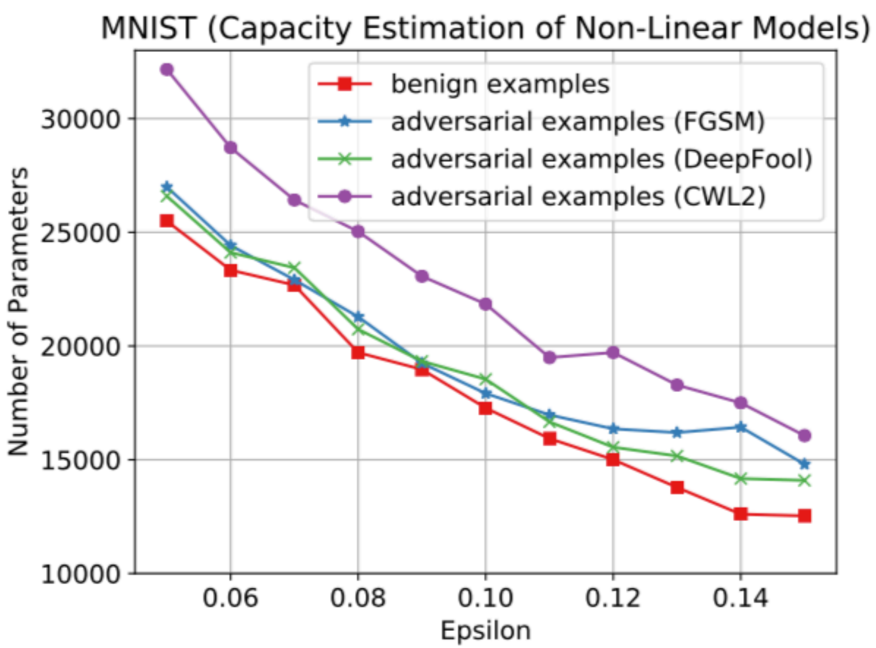
Adversarial Examples always have more noise!

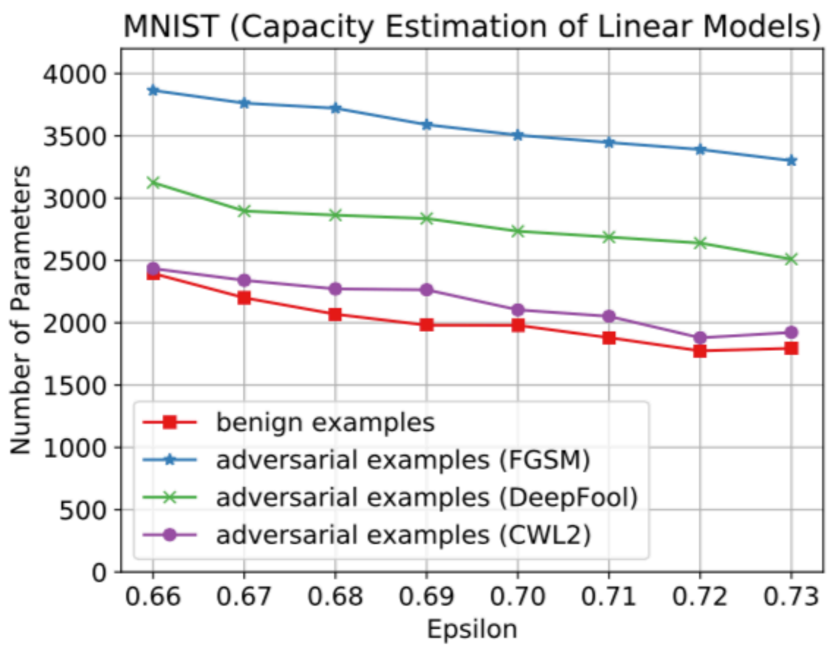
Table 1: Comparison of complexity for benign and adversarial examples on MNIST and CIFAR-10.

Dataset	Examples	H (MLE)	H (JVHW)	Original Size	Compressed Size
MNIST [27]	**Benign**	**1.741**	**1.887**	**988.89 B**	**431.40 B**
	FGSM [16]	2.488	2.601	1690.36 B	503.54 B
	DeepFool [36]	4.844	5.088	1654.99 B	510.41 B
	CW (L_2) [4]	4.094	4.301	1159.01 B	437.27 B
CIFAR-10 [25]	**Benign**	**9.595**	**7.104**	**1845.98 B**	**741.36 B**
	FGSM [16]	9.937	7.710	2717.01 B	872.40 B
	DeepFool [36]	9.675	7.147	1880.41 B	743.02 B
	CW (L_2) [4]	9.621	7.113	1850.54 B	741.56 B

Adversarial Examples are caused by Redundancies: Experiments

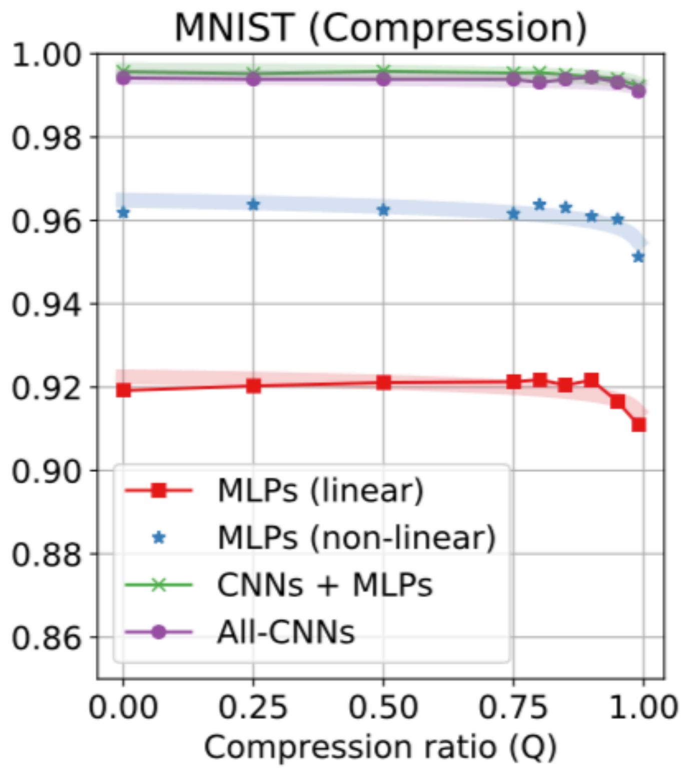
Adversarial Examples always require more capacity to train!





Adversarial avoidance: Reduce Redundancies!

Perceptual compression reduces noise without impacting accuracy!



This will be discussed a lot deeper in the next lecture.

Solution: Kill redundant edges!

1) Redundancies cause adversarial examples.

2) We cannot change the data (other than converting it into "features").

3) The only thing we can do is set irrelevant portions of data to 0. This, is set connections to 0 that are irrelevant — which is the same as not having the connection.

Remember: Occam's Razor is a necessity to restrict number of ways of being able to be contradicted (prevent giving 'weight' to irrelevant information influencing the decision)!

Kill irrelevance: How?

Reduce capacity while training for accuracy!

Memorization preserves redundancies.

For binary classifiers:

$$G = \frac{\#correctly\ classified\ instances}{Memory\ Equivalent\ Capacity}\quad [\frac{bits}{bit}]$$

G<*1* => *M* needs more training/data (not even memorizing)
G=*1* => *M* is memorizing = overfitting
1<G<G_{MEM}=> M could be implementing a lossless compression (and still overfit)
G>G_{MEM}=> *M* is generalizing (no chance for overfitting)

That’s it for today.

Questions?