Experimental Design for Machine Learning on Multimedia Data
Lecture 7

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

Website: http://www.icsi.berkeley.edu/~fractor/fall2019/

Projects

- **Thank you for your submissions**
- **Everybody should have received feedback**
- **No more homework: Please work on your project. Come to office hours for advice!**

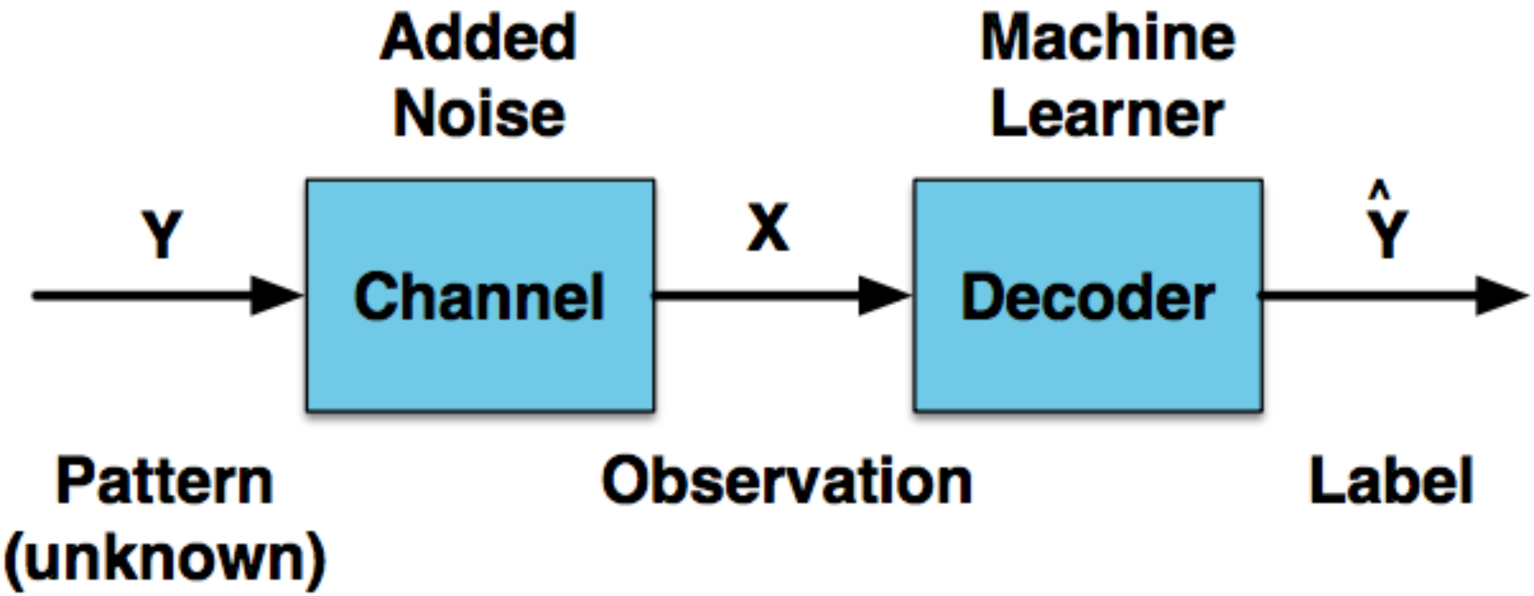
Training I

- **Everything we did so far assumes perfect training. This is, training that guarantees to reach the global minimum error.**
- **Perfect training requires exponential time.**
- **Imperfect training means Memory Equivalent Capacity is effectively reduced.**
- **How to measure that: ?**

My Hypothesis

- **With perfect training and perfect capacity adjustment, machine learning is reproducible.**
- **This is, hyper parameters, initialization, and architecture do not matter!**
- **TODO: Improve training!**

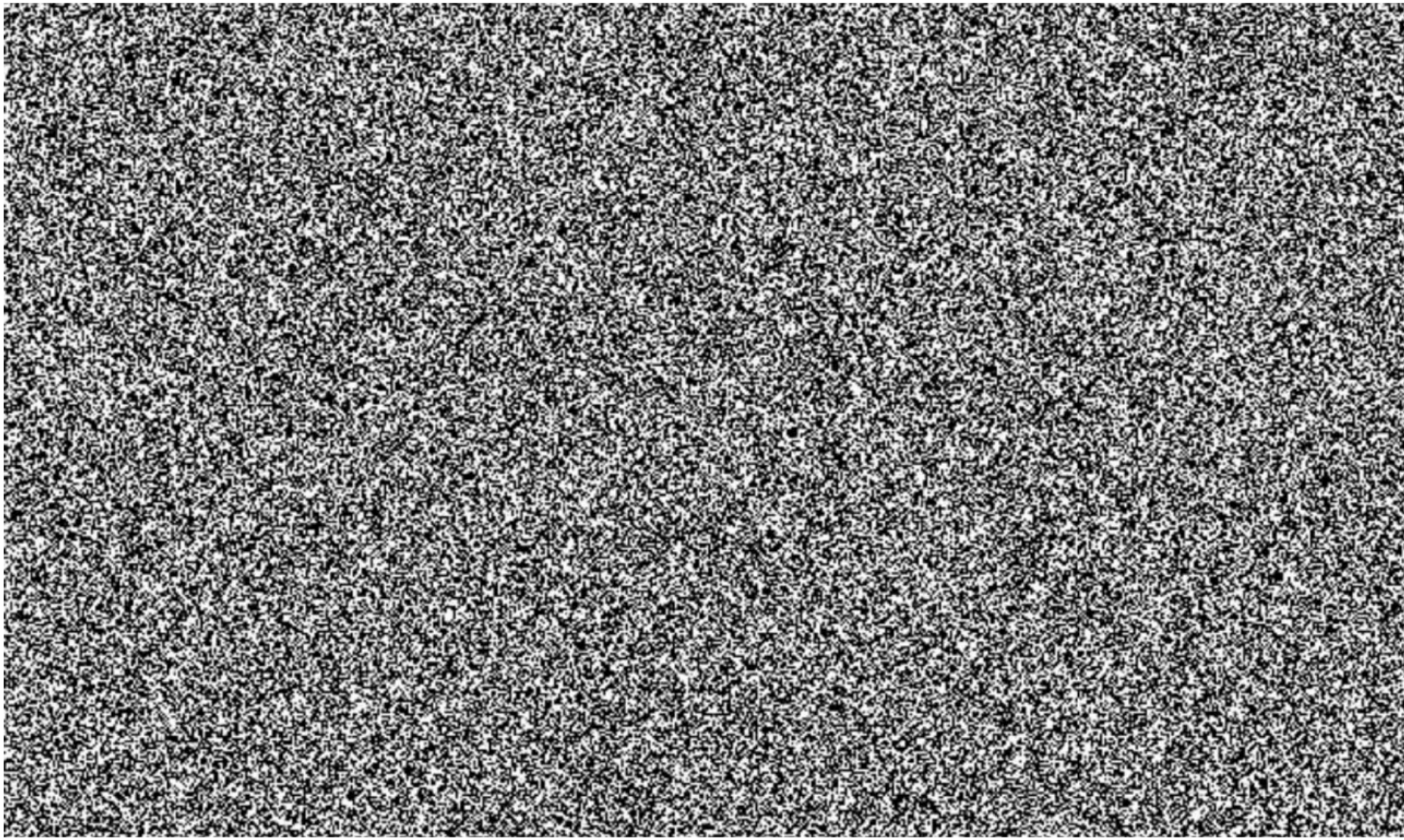
A Thermodynamic/Information Model for ML



- **Machine Learning resets bits introduced by noise.**
- **Machine Learning de*noises* an unknown pattern.**

So: Let's Deal with Noise…

A Video Featuring Nothing But White Noise Has Received Five Content ID Claims Since 2015



Helmholtz free Energy

$$A \equiv U - TS,$$

- A= Free Energy
- U = Internal Energy
- T = Temperature
- S = Uncertainty

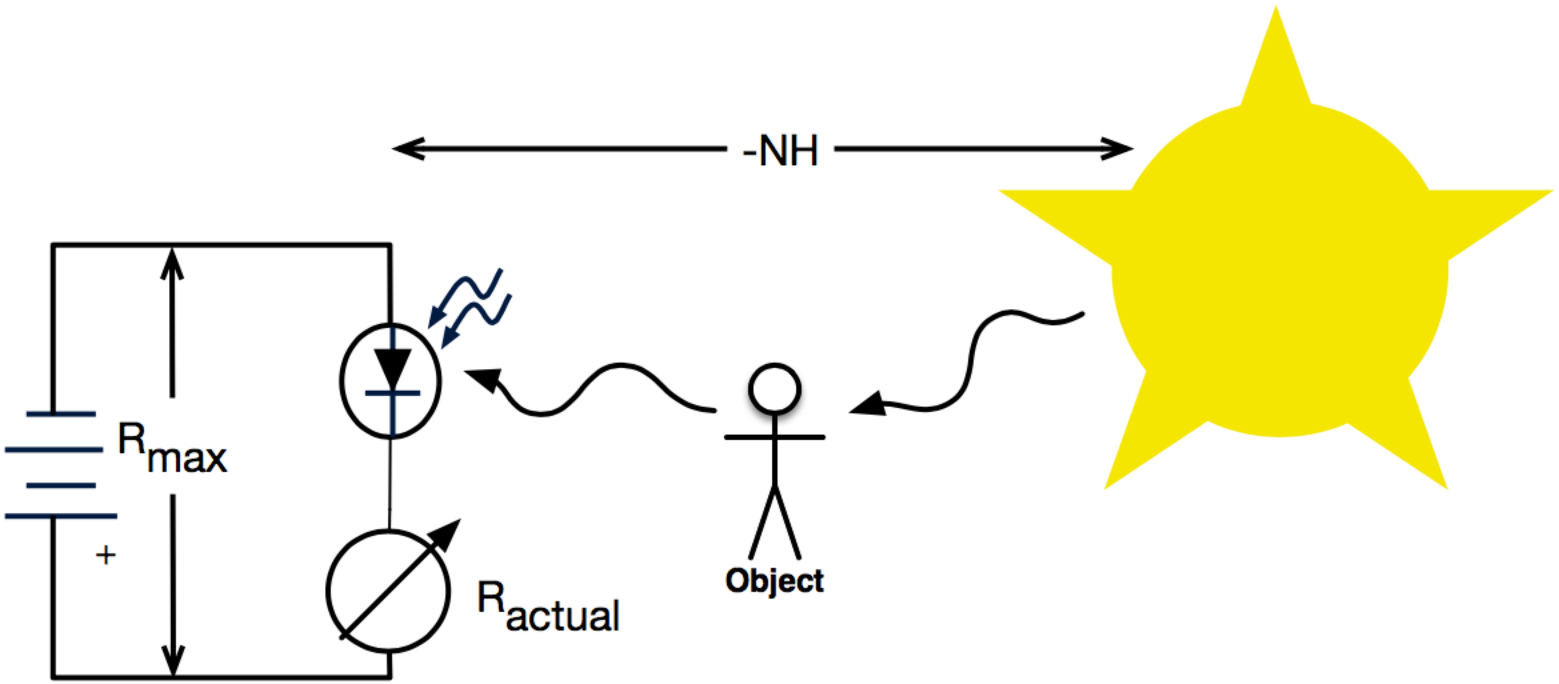
Shannon Entropy and Thermodynamic Entropy

$$H = -S$$

Information is Reduction of Uncertainty

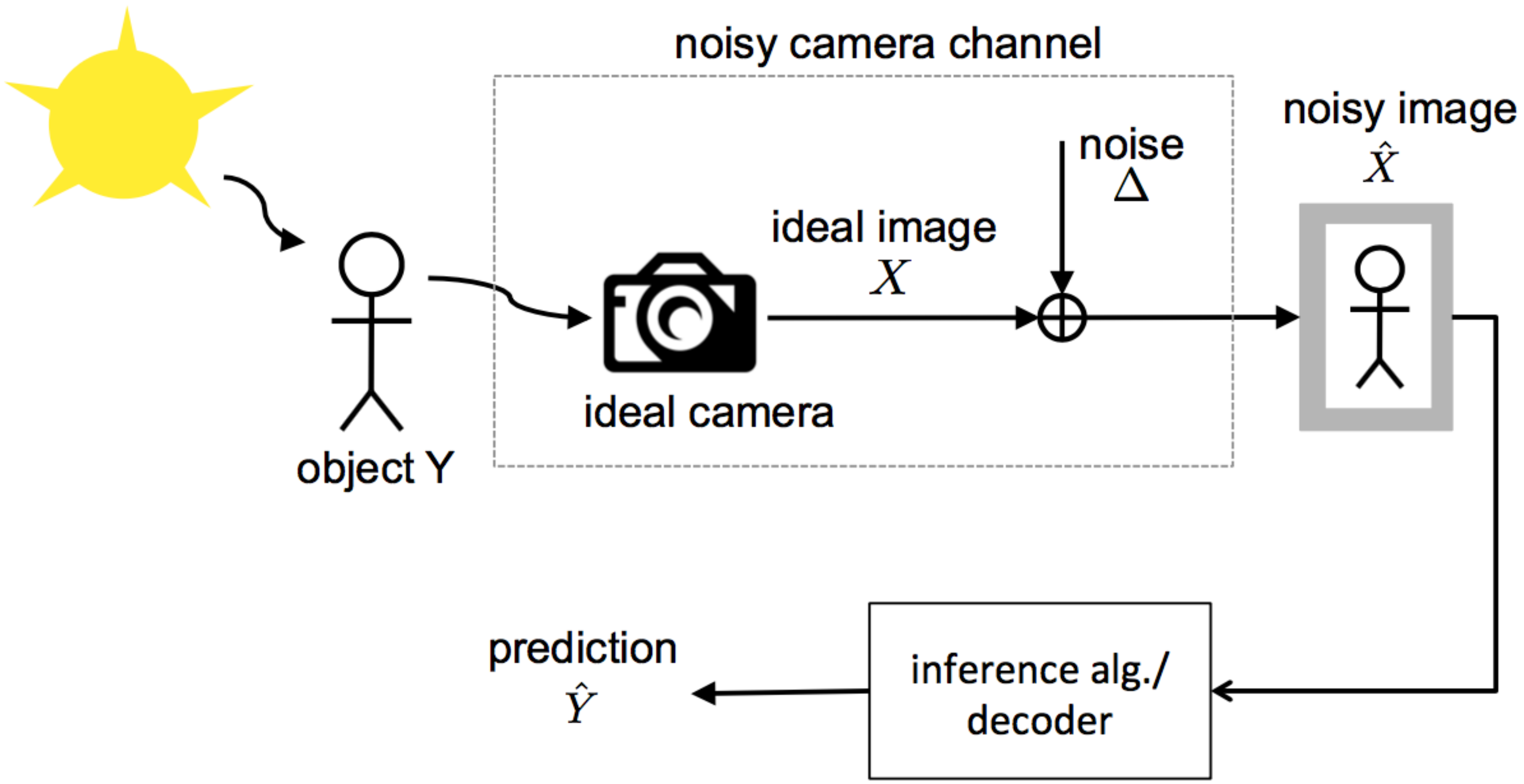
See also: Computation, Data and Science
https://www.youtube.com/playlist?list=PL17CtGMLr0Xz3vNK31TG7mJIzmF78vsFO

Reinterpretation



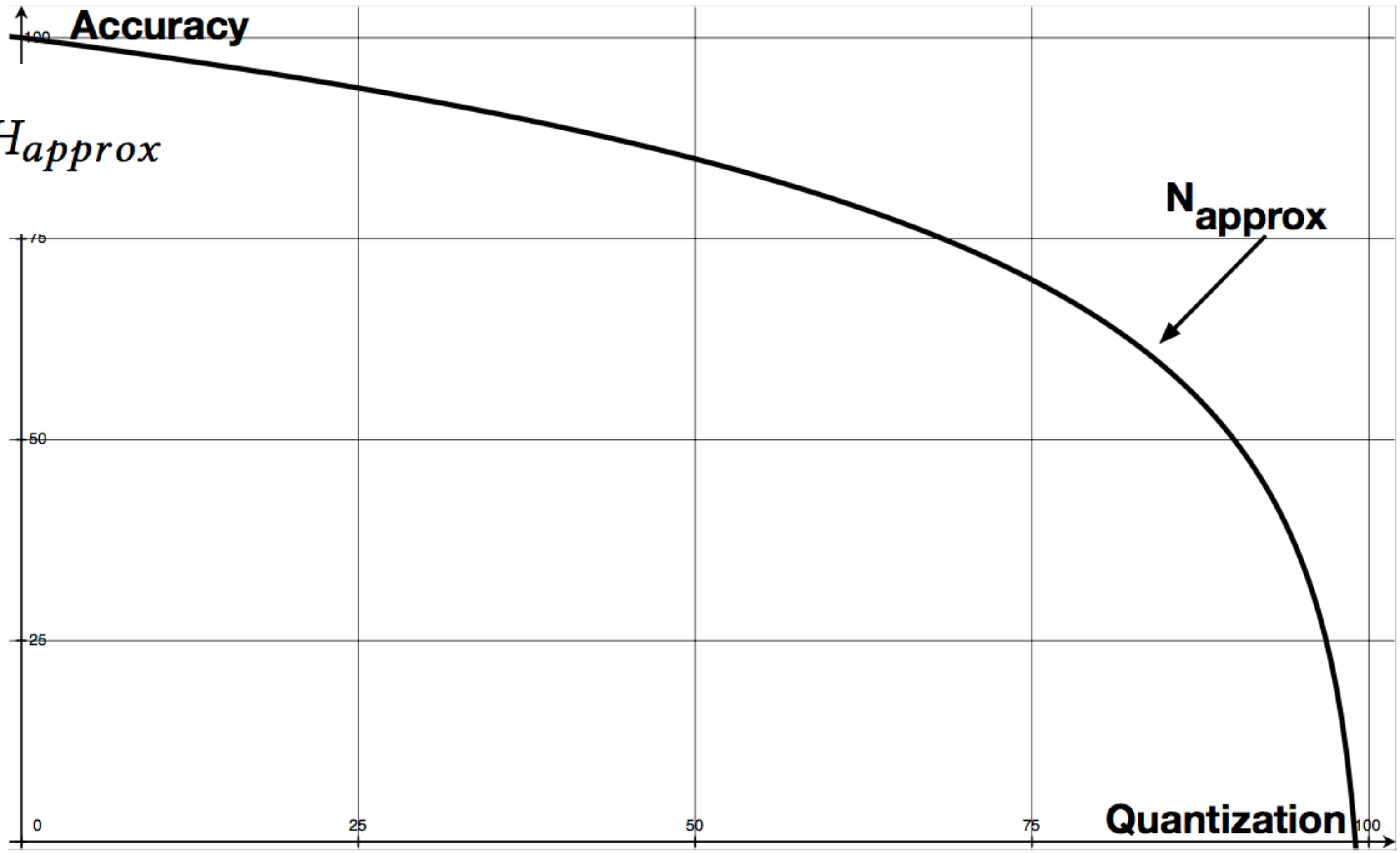
$$R_{actual} = R_{max} - NH$$

Reinterpretation with Information Theory



How does lossy compression work?

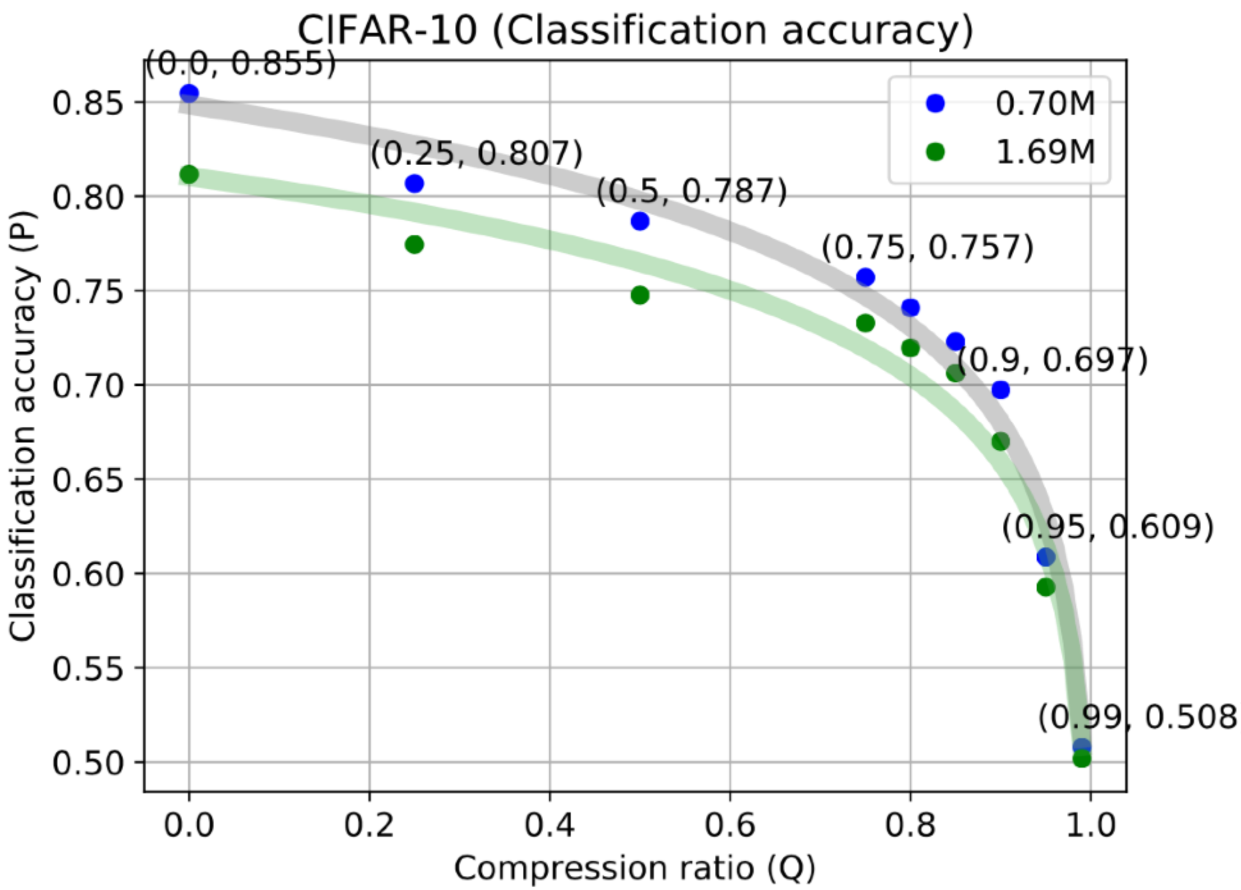
$$\frac{NH}{N_{approx}} = H_{approx}$$



Experiments: Images (overall)

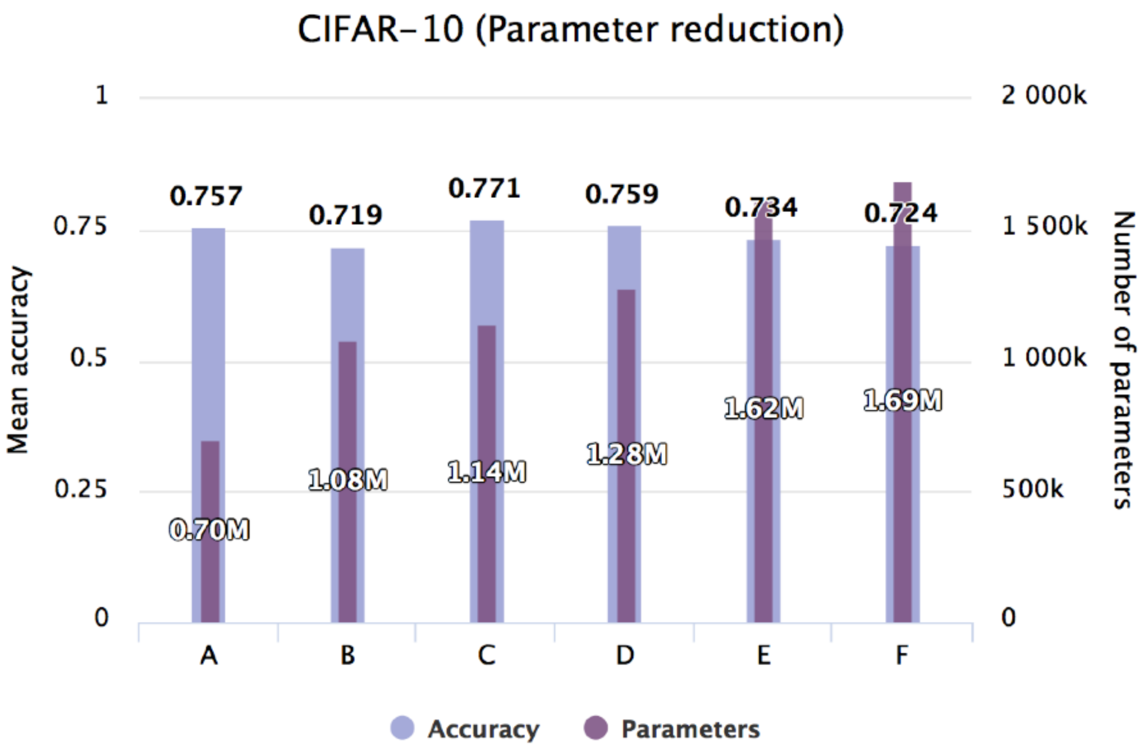
A	C	F
Conv([32, 64], 3, 3) + ReLU	Conv([32, 64], 3, 3) + ReLU	Conv([32, 64], 3, 3) + ReLU
Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)
Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU
Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)	Conv(128, 3, 3) + Dropout(0.5)
Conv([128, 128], 3, 3) + ReLU	Conv([128, 128], 3, 3) + ReLU	Flatten
Conv(10, 3, 3)	Conv(128, 3, 3) + Dropout(0.5)	FC(128) + Dropout(0.5)
Global_avg_pooling	Conv([128, 128], 3, 3) + ReLU	FC(256) + Dropout(0.5)
Softmax	Conv(10, 3, 3)	FC(256) + Dropout(0.5)
	Global_avg_pooling	FC(10)
	Softmax	Softmax
701,386 (0.70M)	1,144,138 (1.14M)	1,686,090 (1.69M)

Experiments: Images (overall)



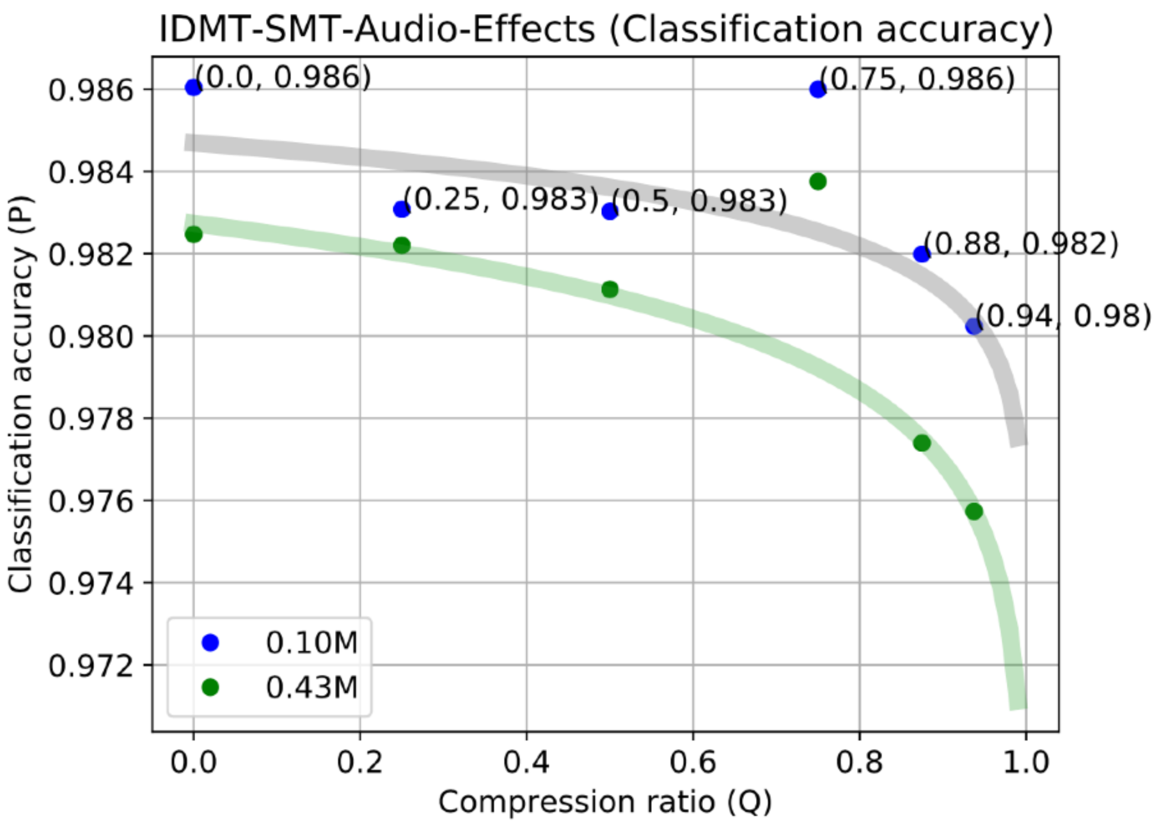
No Free Lunch!

Experiments: Images concrete



Less Parameters = Higher Accuracy!

Experiments: Audio



Experiments generalize to audio

Analysis: Images

JPEG quantization matrizes:

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	36	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Best quality/accuracy trade-off (N_{approx}) around q=20.
This is at 1 bit/pixel!

Sources

Jingkang Wang, Ruoxi Jia, Gerald Friedland, Bo Li, Costas Spanos:
One Bit Matters: Understanding Adversarial Examples as the Abuse of Redundancy, https://arxiv.org/abs/1810.09650

Gerald Friedland, Jingkang Wang, Ruoxi Jia, Bo Li:
The Helmholtz Method: Using Perceptual Compression to Reduce Machine Learning Complexity, https://arxiv.org/abs/1807.10569

That's it for today.

Questions?