Microsoft

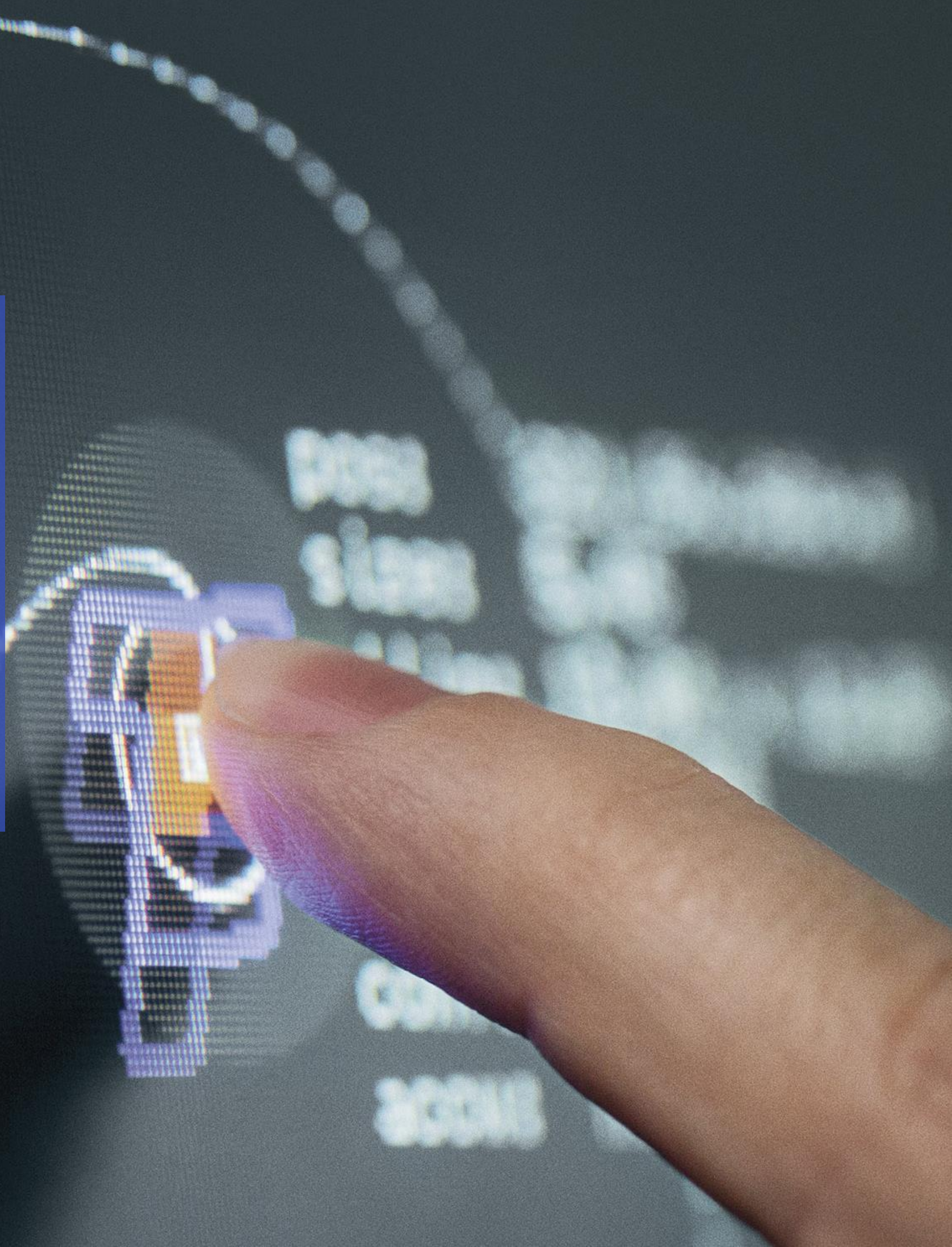
Machine Learning applied to device scenarios

Alisson Sol

Development Lead – Bing Engineering Fundamentals

Think forward

Device Summit 2013



Agenda

- Introduction
- Some ML definitions
- Applying ML to a real problem
- Q&A

My coding for devices

Introduction

Many teams contributed to results shown here:
Kinect for Windows, Microsoft Research, Xbox, etc.

This presentation will focus on practical aspects

Learning ML theory

Paper

A Few Useful Things to Know about Machine Learning.
Communications of the ACM, 55 (10), 78-87, 2012.

Books

Learning From Data, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin

Machine Learning in Action, Peter Harrington

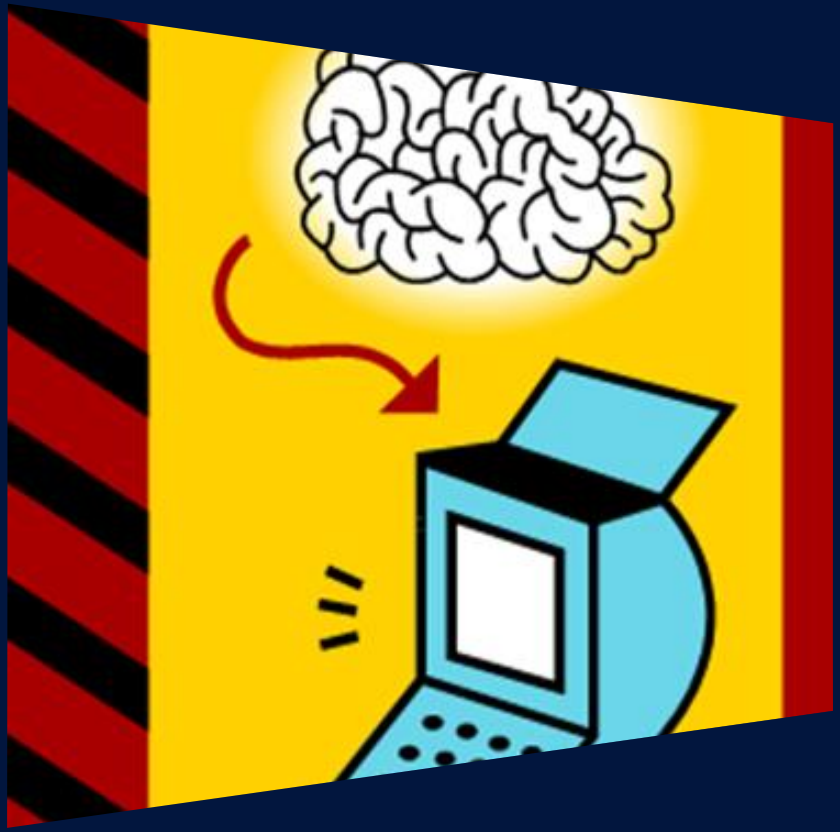
Pattern Recognition and Machine Learning, Christopher M. Bishop

Online courses

https://www.coursera.org/course/machlearning

https://www.coursera.org/course/ml

Key messages



- Data isn't the same as information
- Testing ML requires domain knowledge

Definitions for the problem to be solved

- Pose
- Gesture
- Interaction

Interactions demo

Machine learning in Kinect for Windows

- Skeletal tracking
- Hand events (Grip, GripRelease)

Usual development ML development

ML applications

- Classification
- Grouping
- Ranking
- …

ML definitions

- Feature
- Training
- Verification
- Validation

An example from real estate

ZIP	Beds	Baths	Sq ft	Built	Lot	Update	Value
98052	4	3	2,200	1968	10,000	-	400,000
98004	4	3	2,100	1968	10,000	-	500,000
98004	5	3	2,400	1970	9,000	2005	600,000
98008	4	2.5	2,200	1980	5,000	-	500,000
…	…	…	…	…	…	…	…
98052	4	2	2,200	1990	7,000	-	???

Dev verification x QA validation

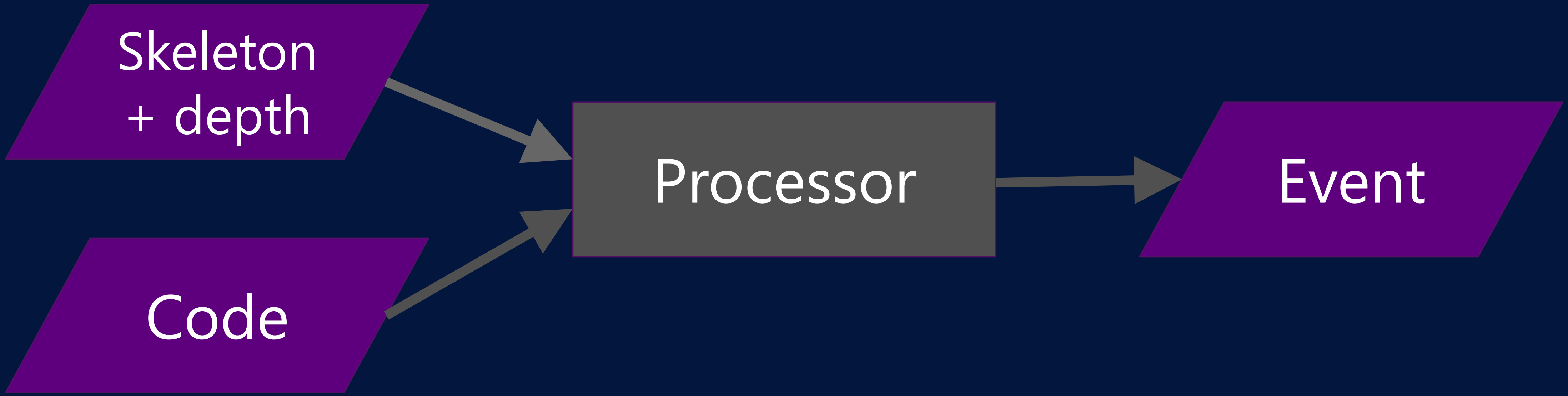
An example from medicine

Age	Tobacco	UV light	Family	Alcohol	BMI	Diagnostic
39	0	3	2	0	20.5	0
45	5	3	0	0	21.3	0
52	15	5	0	3	29	1
68	20	3	1	7	23	0
…	…	…	…	…		…
55	5	4	1	9	25	???

Ground truth

Recognizer

ML-based recognizer



Key messages

- Data isn't the same as information
- Testing ML requires domain knowledge

Problem size

Features

Skeleton data, deltas (among joints), time deltas (across frames), depth differences (in frame, over time), etc…

Depth data per hour of recording

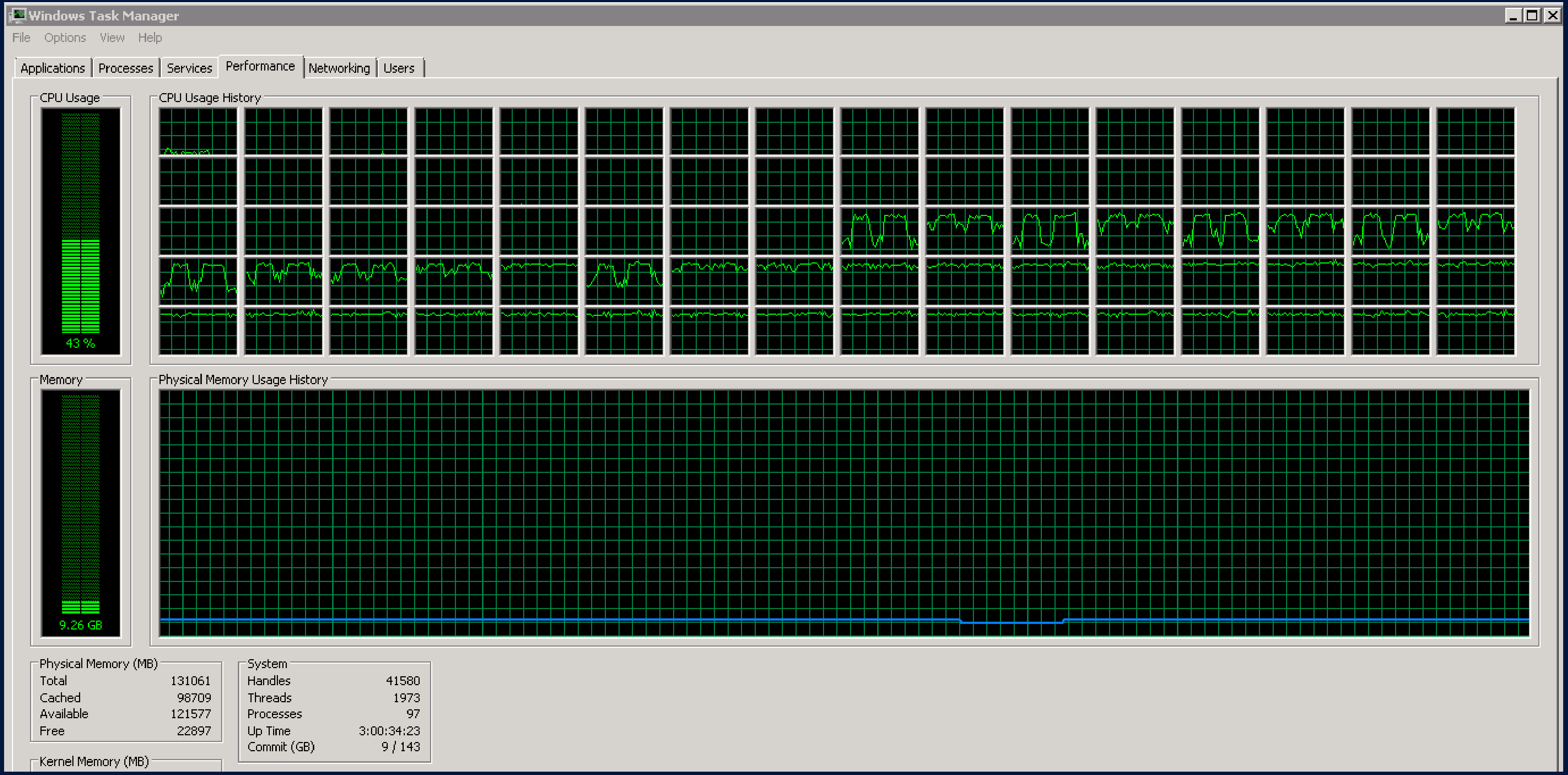
60 min/hour x 60 sec/min x 30 fps =~ 108,000 frames

Each frame = 640 x 480 = 300 K pixels

Section around a hand = 128 x 128 = 16 K pixels

Frame depth differences =~ 16 K x 16 K x 0.5 =~ 128 M features

It is a costly computational problem...



Selecting a training method

A Few Useful Things to Know about Machine Learning

Pedro Domingos
Department of Computer Science and Engineering
University of Washington
Seattle, WA 98195-2350, U.S.A.
pedrod@cs.washington.edu

ABSTRACT

Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. This is often feasible and cost-effective where manual programming is not. As more data becomes available, more ambitious problems can be tackled. As a result, machine learning is widely used in computer science and other fields. However, developing successful machine learning applications requires a substantial amount of "black art" that is hard to find in textbooks. This article summarizes twelve key lessons that machine learning researchers and practitioners have learned. These include pitfalls to avoid, important issues to focus on, and answers to common questions.

correct output y_t for future examples $\mathbf{x_t}$ (e.g., whether the spam filter correctly classifies previously unseen emails as spam or not spam).

2. LEARNING = REPRESENTATION + EVALUATION + OPTIMIZATION

Suppose you have an application that you think machine learning might be good for. The first problem facing you is the bewildering variety of learning algorithms available. Which one to use? There are literally thousands available, and hundreds more are published each year. The key to not getting lost in this huge space is to realize that it consists of combinations of just three components. The components

Right method == correctly implemented

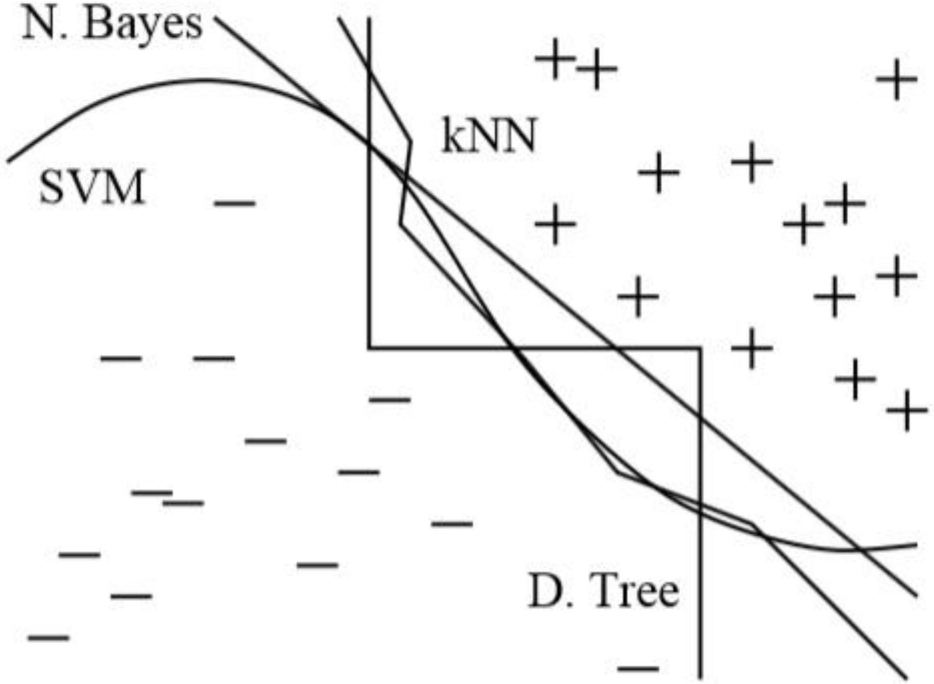
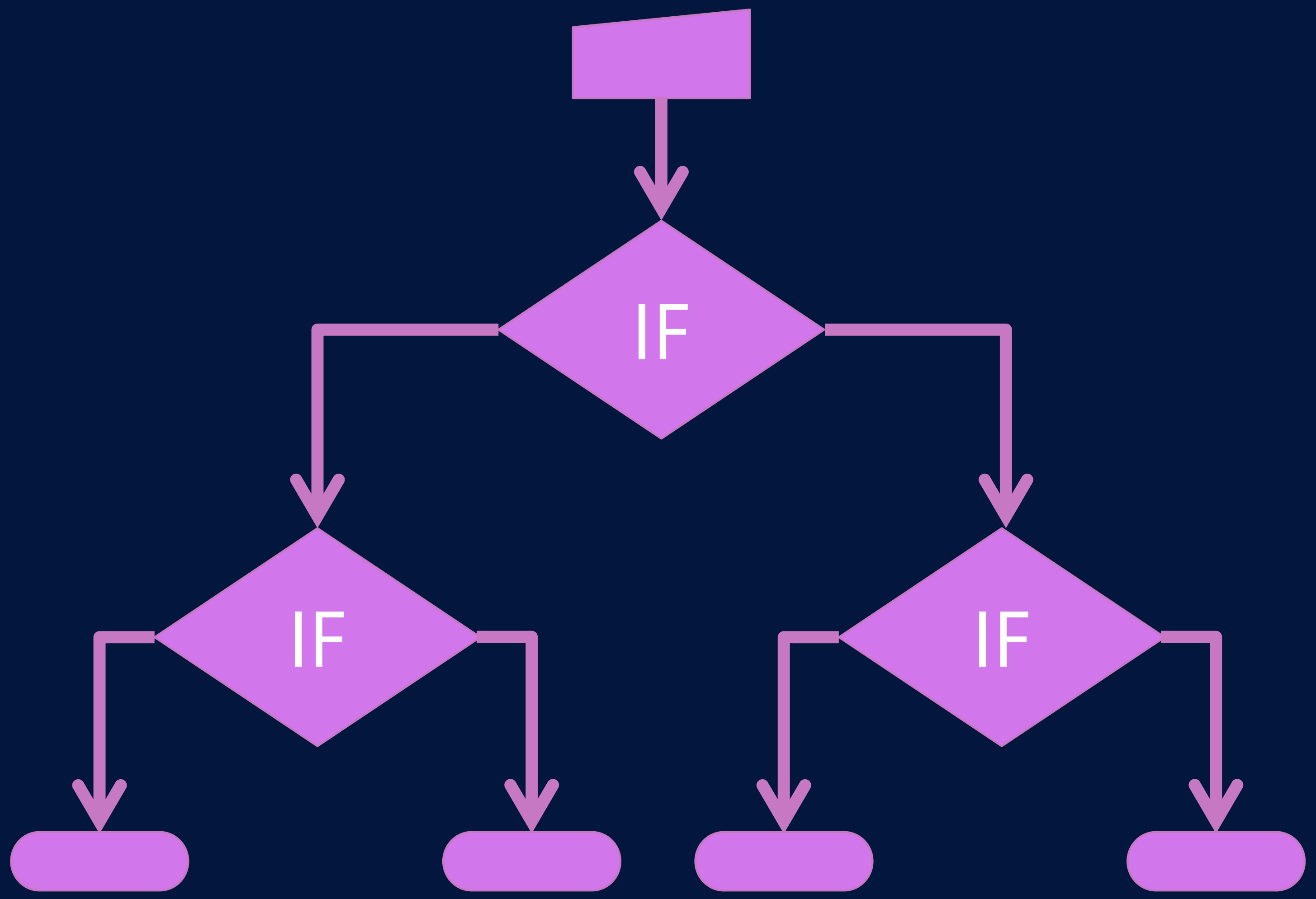


Figure 3: Very different frontiers can yield similar class predictions. (+ and – are training examples of two classes.)

A decision tree



ID3: Iterative Dichotomiser 3

For unused features

Choose feature of maximum information gain

Make node splitting data based on test against that feature

End conditions

Tree depth, lack of data, lack of features

Marine animal data

Survive underwater?	Has flippers?	Fish?
Y	Y	Y
Y	Y	Y
Y	N	N
N	Y	N
N	Y	N

Source: book Machine Learning in Action

Which feature to test first?

Information as function of probability for outcomes

$$I(x_i) = \log_2 p(x_i)$$

Shannon's entropy

$$H = -\sum_{i=1}^{n} p(x_i) I(x_i)$$

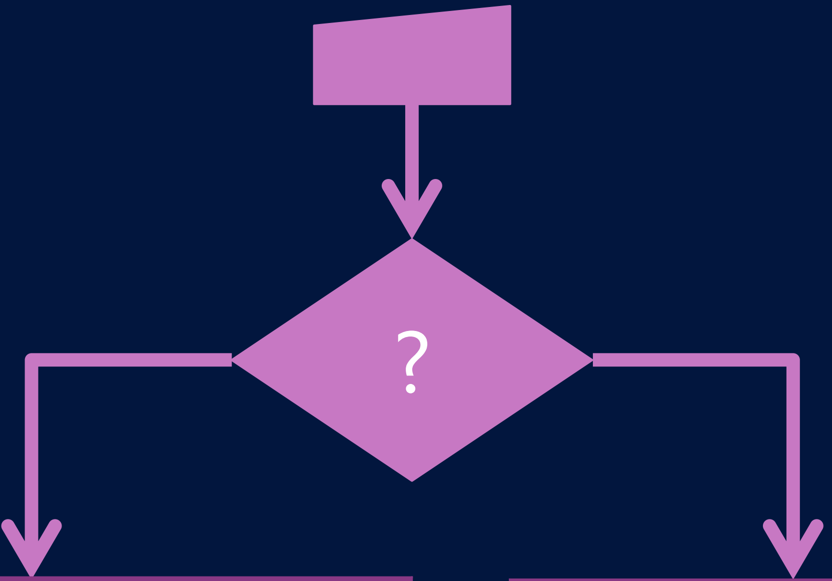
Calculate entropy

Under?	Flipper?	Fish?	Prob.	Log2(Prob.)	Entropy
Y	Y	Y			
Y	Y	Y			
Y	N	N			
N	Y	N			
N	Y	N			

Applying formulas

Under?	Flipper?	Fish?	Prob.	Log2(Prob.)	Entropy
Y	Y	Y	0.4	-1.32	0.53
Y	Y	Y			
Y	N	N	0.6	-0.74	0.44
N	Y	N			
N	Y	N			
					0.97

"Flippers?"

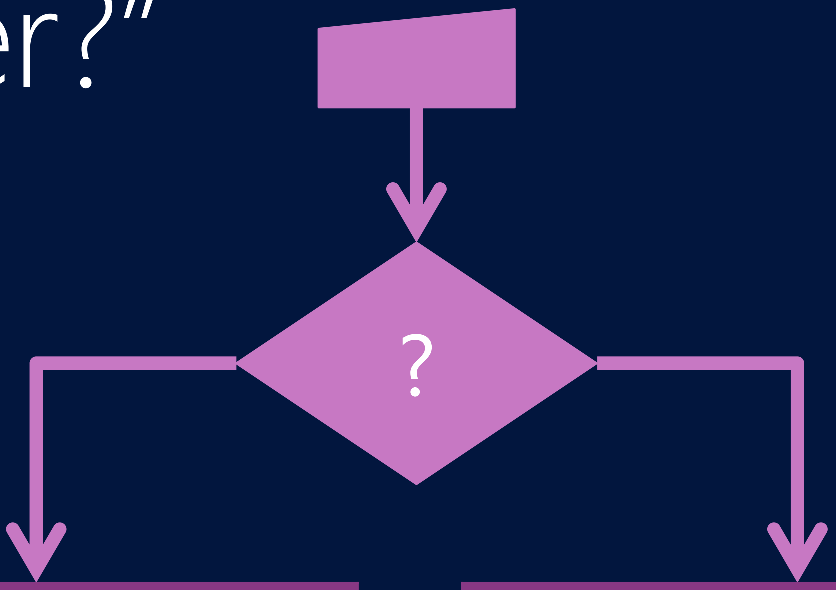


Under?	Flipper?	Fish?	P.	Log2(P.)	Entropy
Y	Y	Y	0.5	-1.00	0.50
Y	Y	Y			
N	Y	N	0.5	-1.00	0.50
N	Y	N			
					1.00

Under?	Flipper?	Fish?	P.	Log2(P.)	Entropy
Y	N	N	1	0.00	0.00

H(Flippers?) = H(Y)*4/5 + H(N)*1/5 = 1.0*4/5 + 0*1/5 = 0.8

"Underwater?"

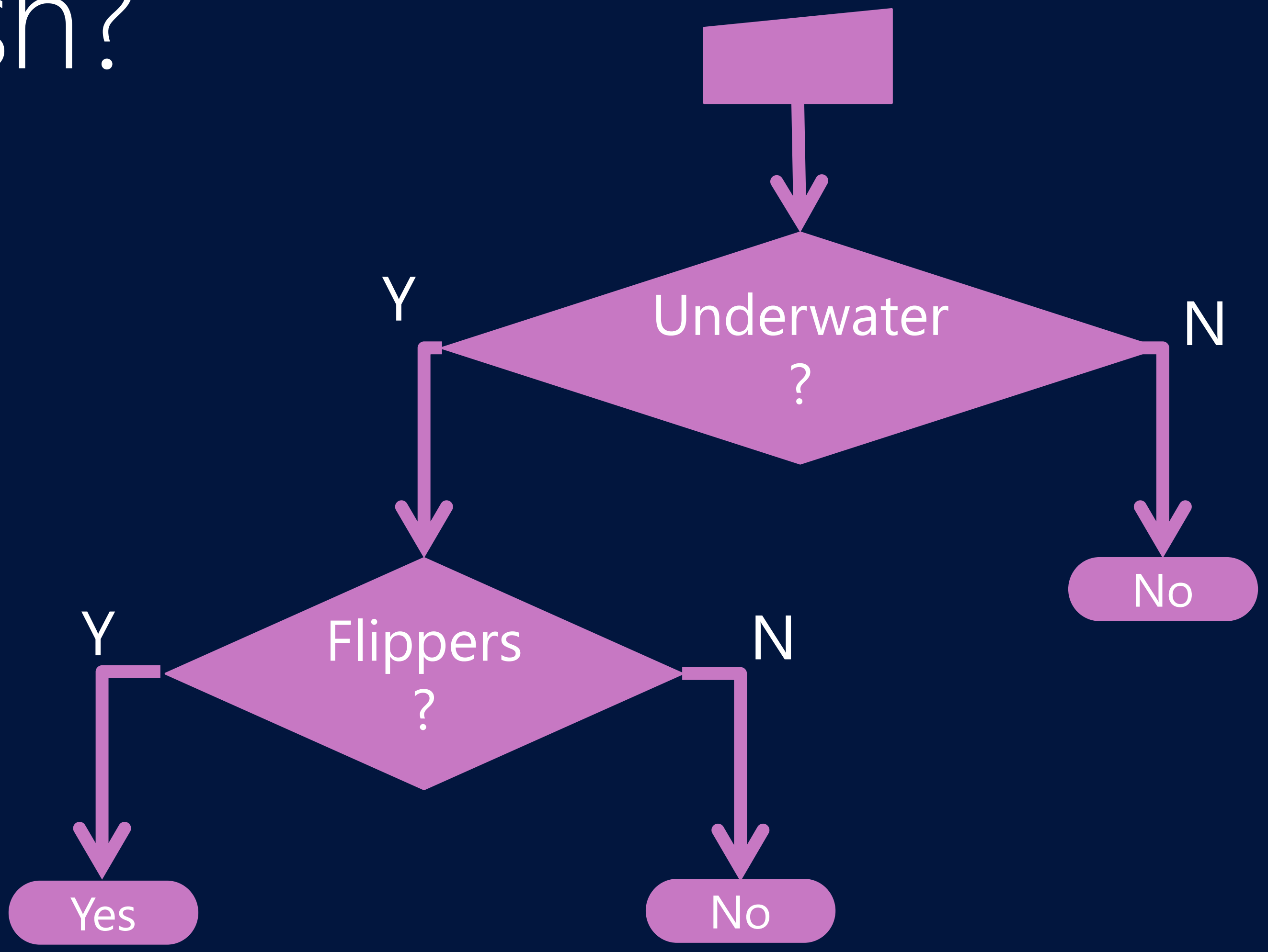


Under?	Flipper?	Fish?	P.	Log2(P.)	H
Y	Y	Y	0.67	-0.58	0.39
Y	Y	Y			
Y	N	N	0.33	-1.58	0.53
					0.92

Under?	Flipper?	Fish?	P.	Log2(P.)	Entropy
N	Y	N	1	0.00	0.00
N	Y	N			

$$H(\text{Underwater?}) = H(Y)*3/5 + H(N)*2/5 = 0.92*3/5 + 0*2/5 = 0.552$$

Is Fish?

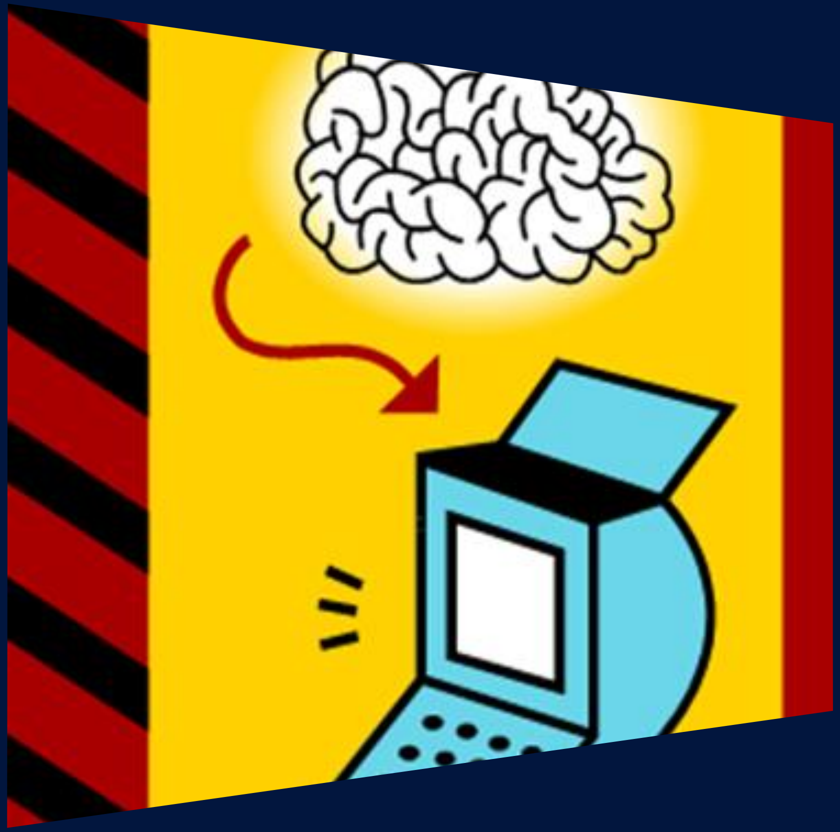


End of first fast road…

Ommitted several details

- Labels may not be boolean
- “Ground truth” can have wrong labels
- …
- Your resulting model may not predict anything!
  - Because of the data…
  - Because of the model…

Key messages



- Data isn't the same as information
- Testing ML requires domain knowledge

Why a ML classifier may fail (in production)?

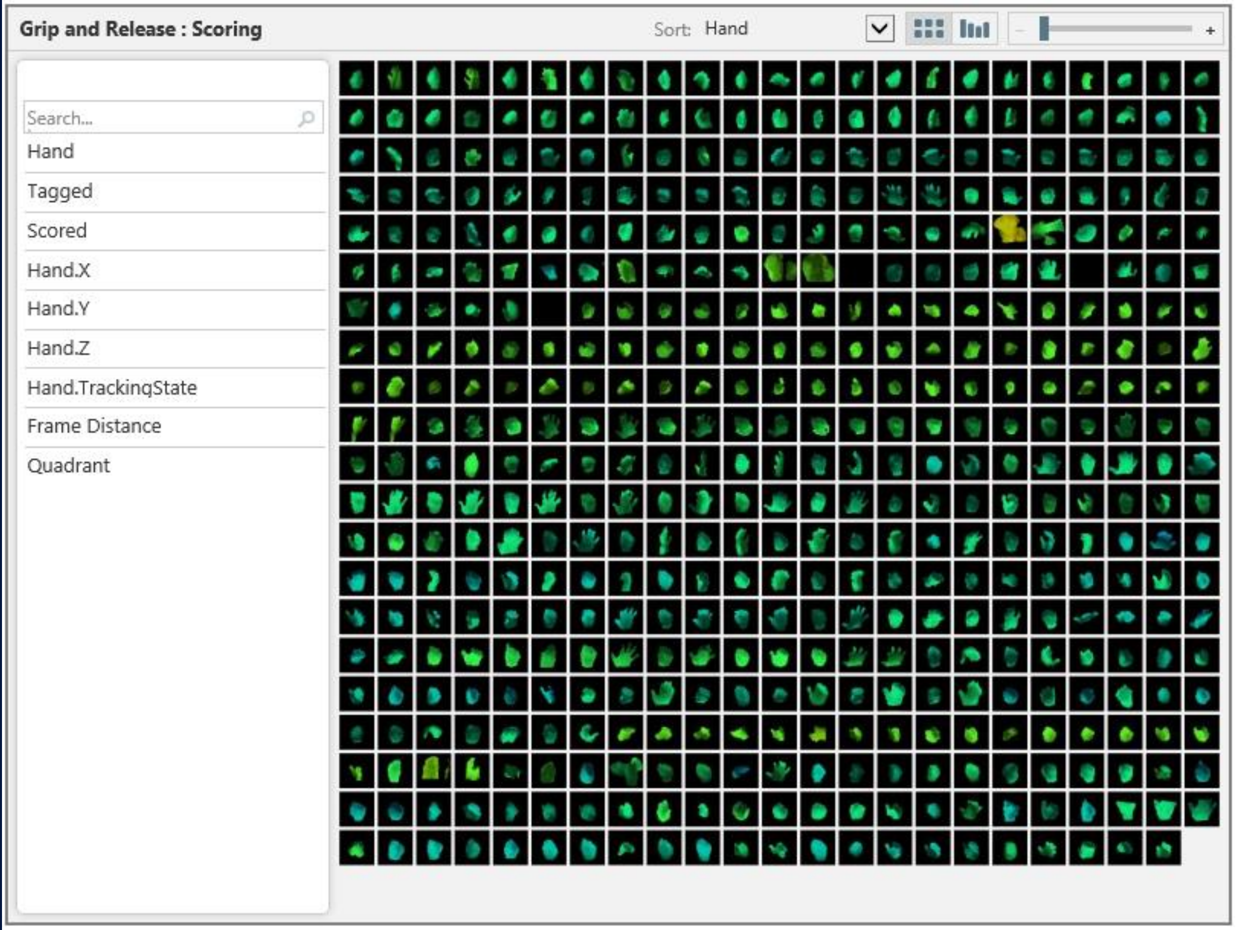
- Wrong features
- "Hidden/unknown" features
- Wrong implementation
  - Debugging ML-based systems is hard
- Wrong testing…

Test data and scenarios…

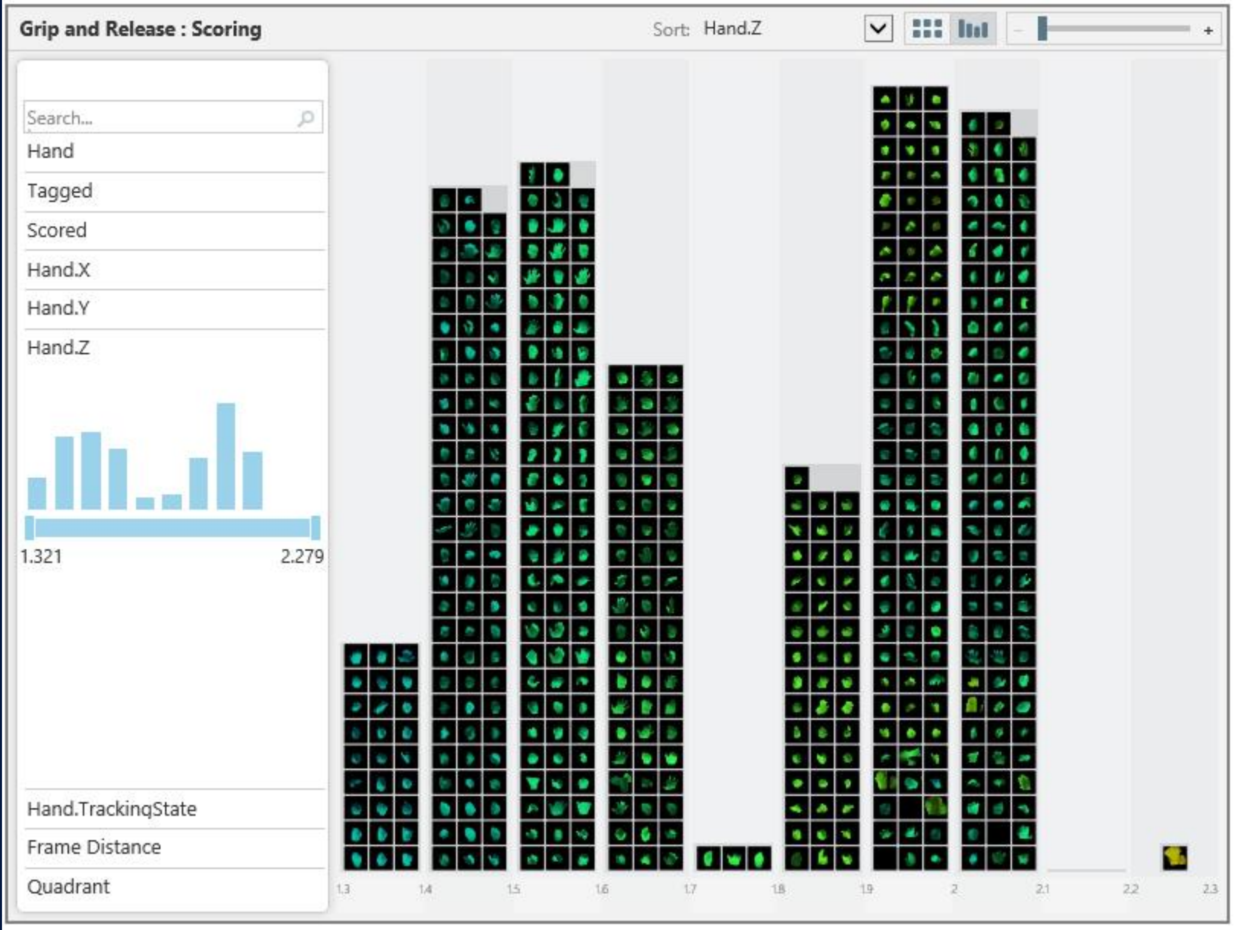
May not represent entire “population”

May not be large enough for extrapolation of results

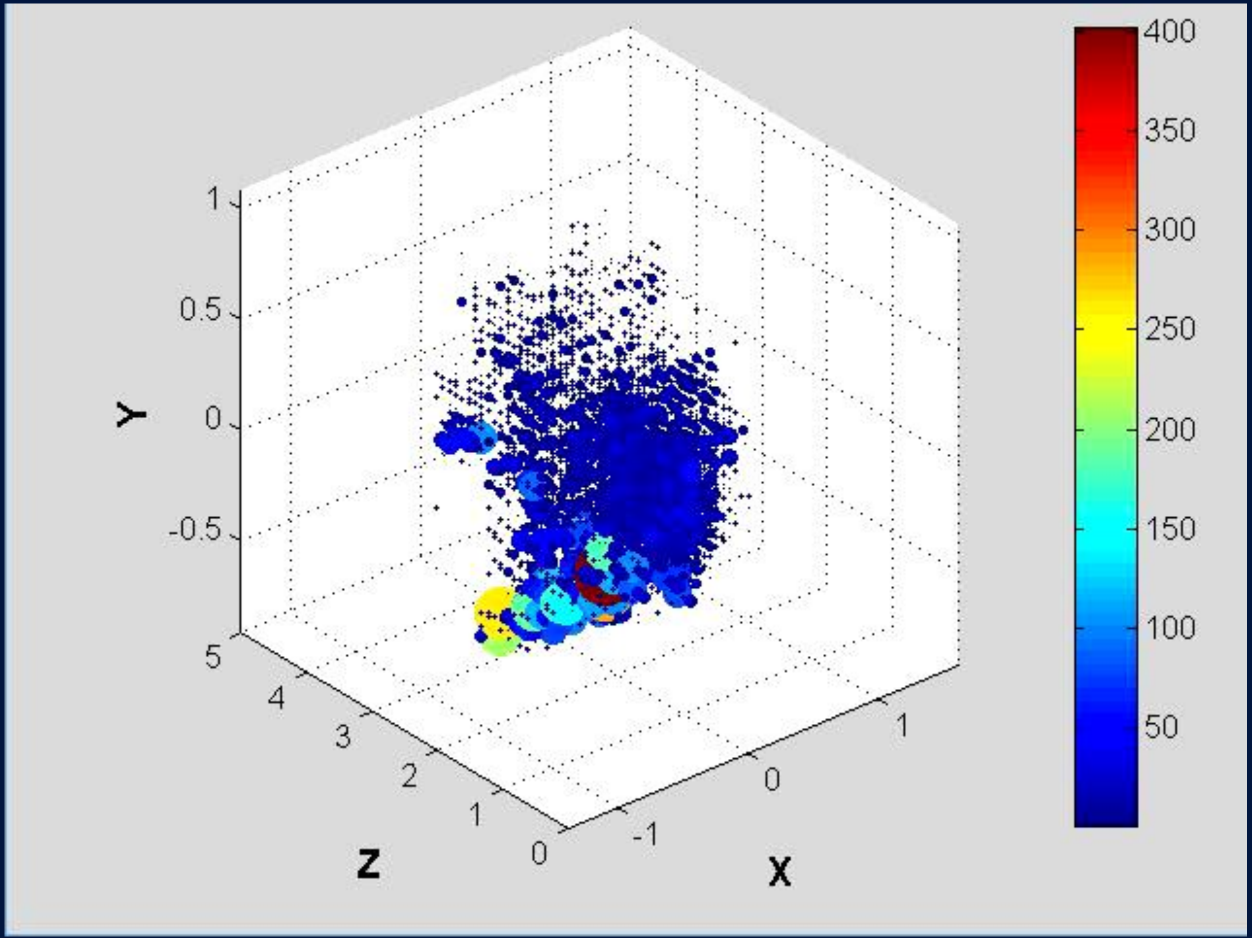
Data analysis for an experiment



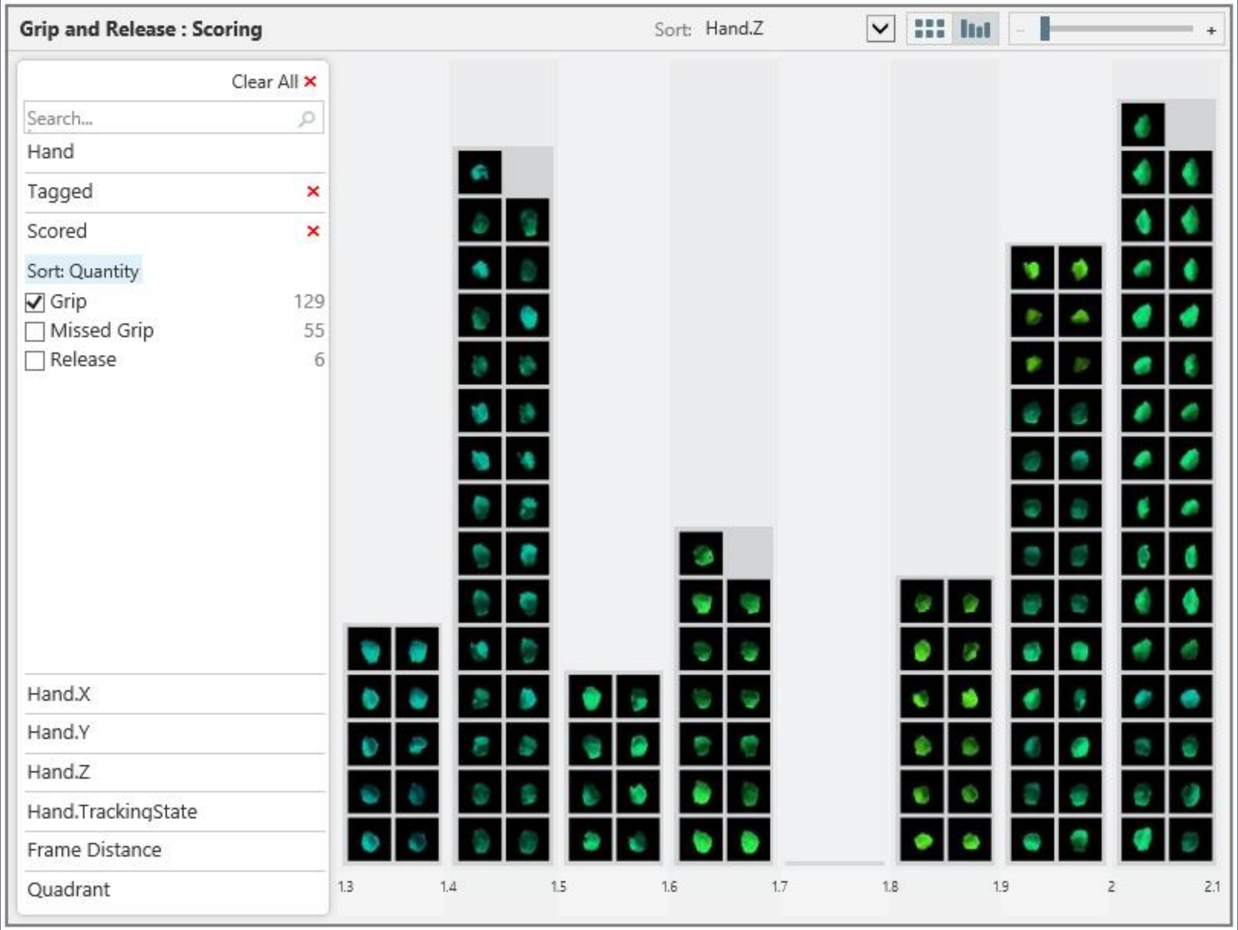
Sort by distance



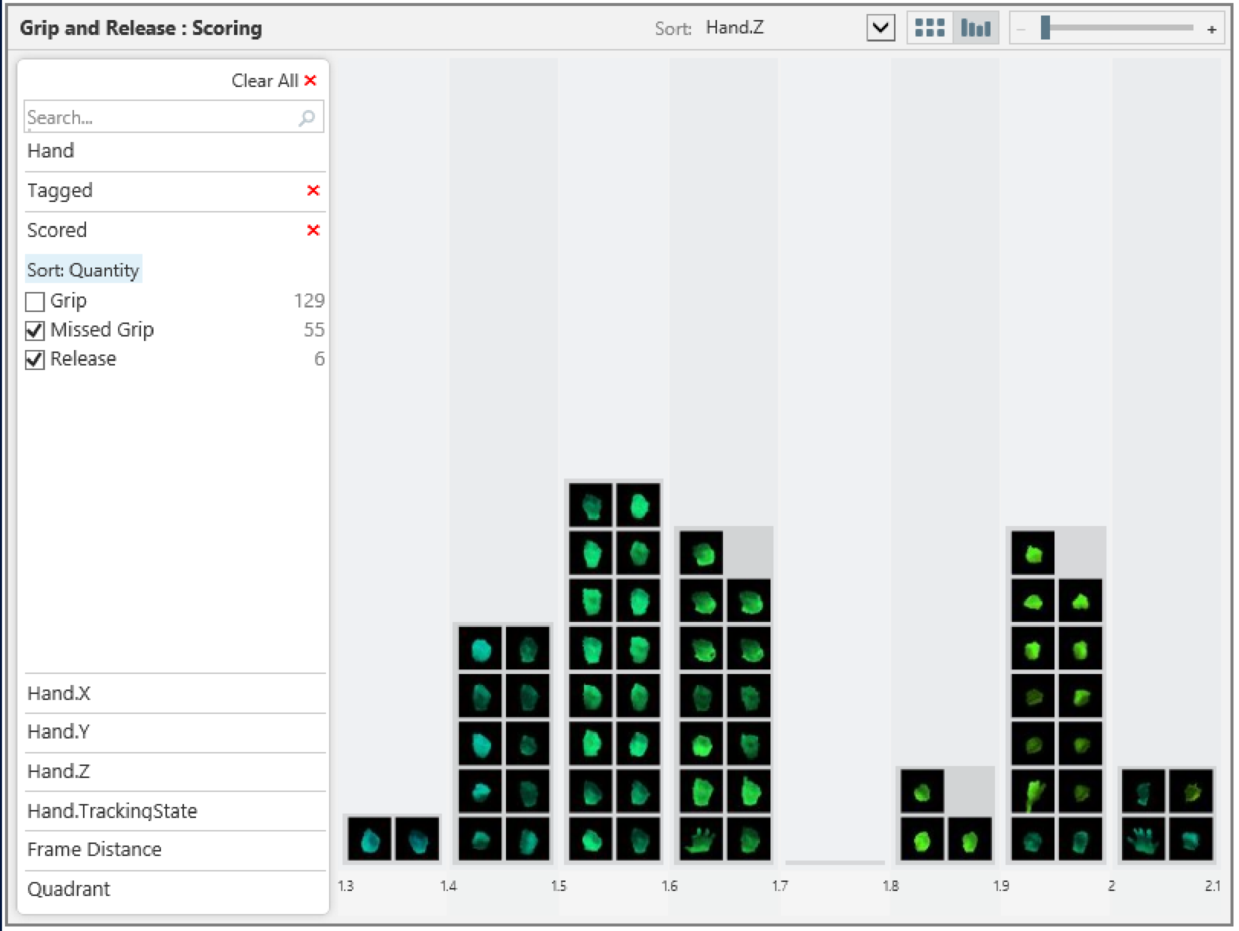
Data in 3-D space



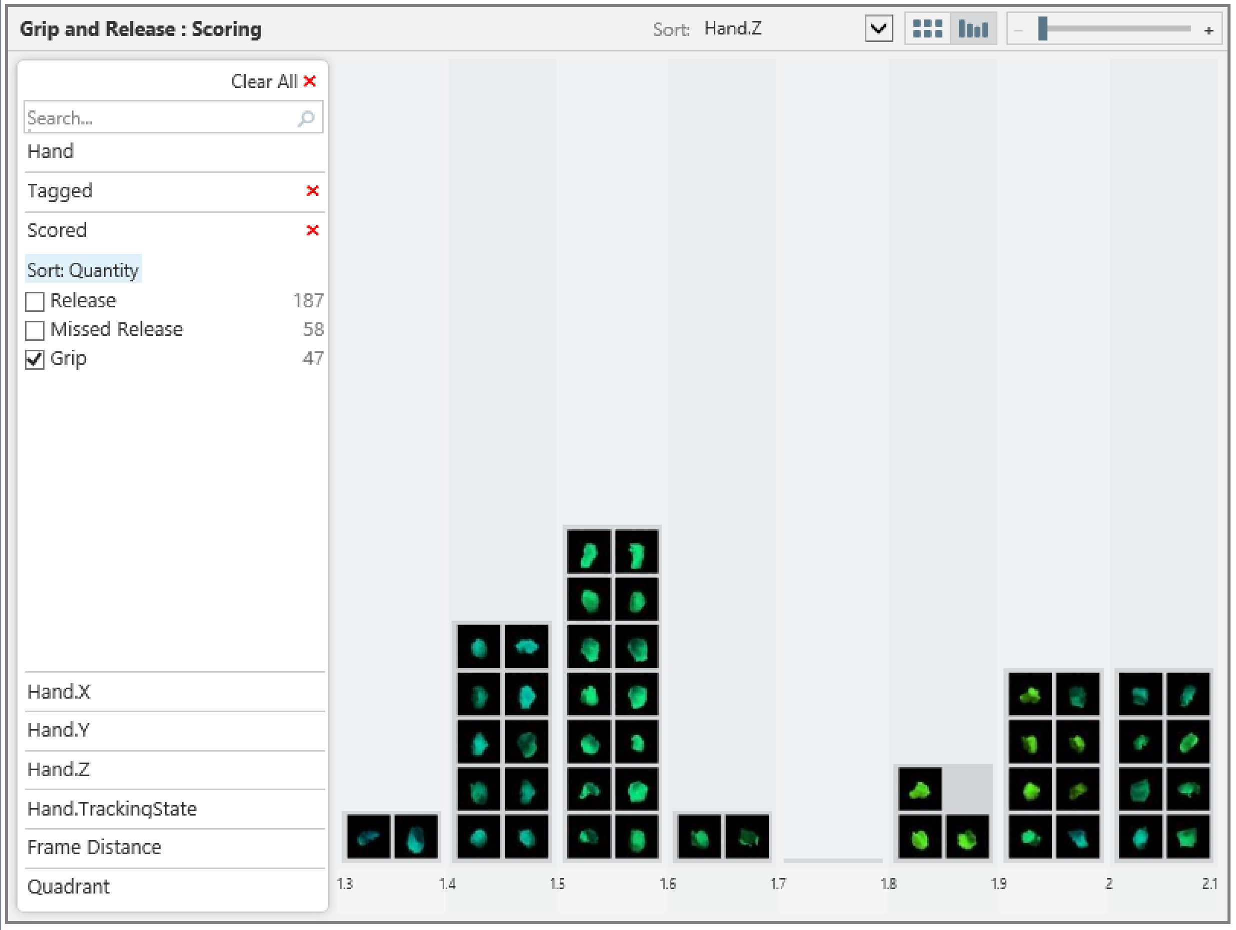
Grip true positives



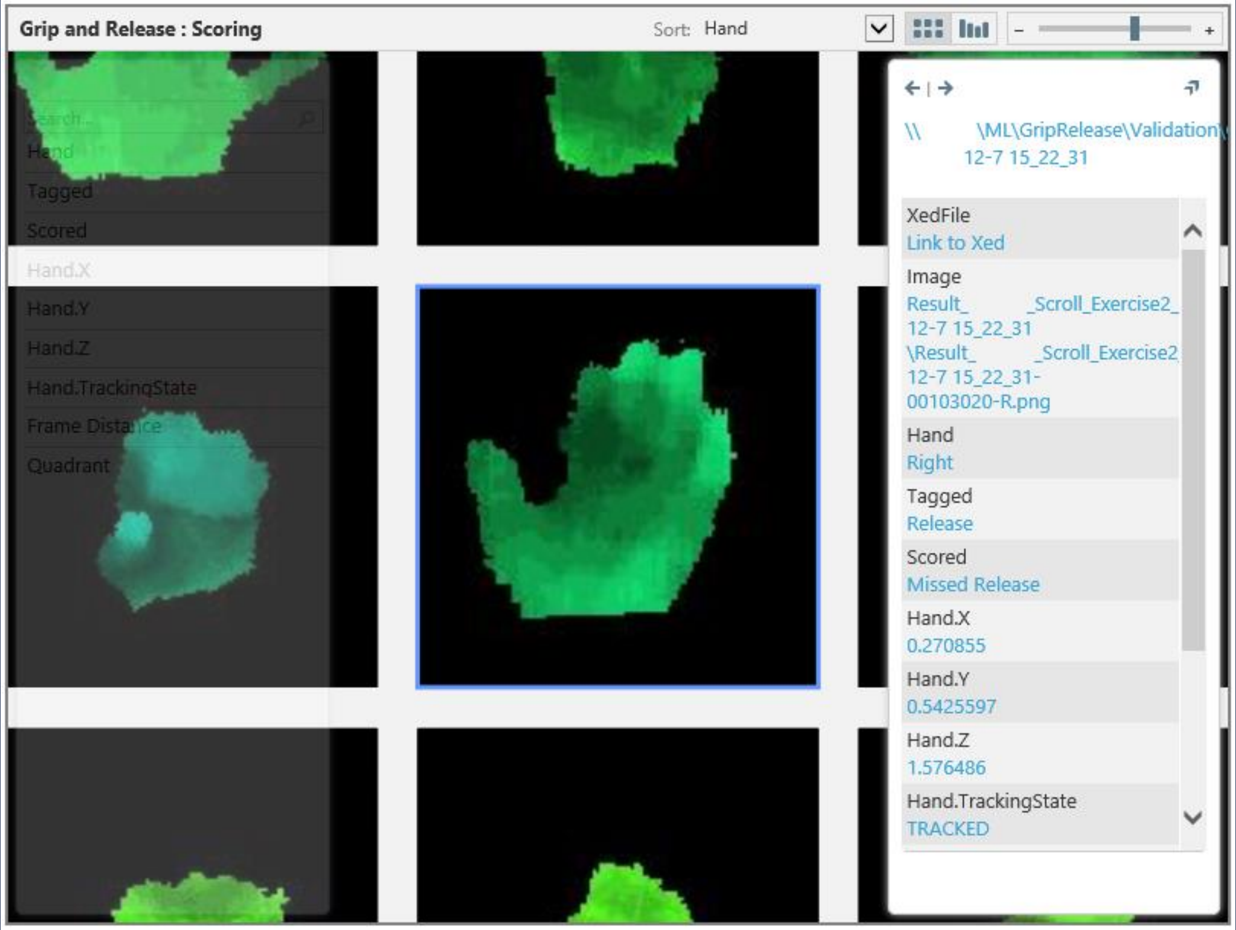
Grip false negatives



Grip false positives



Drill-down to details



Data and metrics

- True positives, false negatives, false positives
- Overwhelming true negatives influence on metrics
  - Precision = TP / (TP + FP)
  - Recall = TP / (TP + FN)
  - Accuracy = (TP + TN) / (TP + TN + FP + FN)
  - F-Score = 2 * (Precision * Recall) / (Precision + Recall)

Warning

The following “theory-related” section will be presented intentionally fast…

Second and final time we browse through theory…

Sample size for extrapolation of results

$$n = \frac{\chi^2 * N * P * (1 - P)}{\left(\delta^2 * (N - 1)\right) + (\chi^2 * P * (1 - P))}$$

N = population size

P = proportion of population with expected feature

χ^2 = Chi-square for the confidence of 1st degree of freedom

δ = target error margin

n = sample size

Source: Determining Sample Size for Research Activities
Robert V. Krejcie and Daryle W. Morgan
Educational and Psychological Measurement, Vol 30, pp. 607-610, 1970

Reusing example from real estate

ZIP	Beds	Baths	Sq ft	Built	Lot	Update	Value
98052	4	3	2,200	1968	10,000	-	400,000
98004	4	3	2,100	1968	10,000	-	500,000
98004	5	3	2,400	1970	9,000	2005	600,000
98008	4	2.5	2,200	1980	5,000	-	500,000
…	…	…	…	…	…	…	…
98052	4	2	2,200	1990	7,000	-	???

Not that confident to start with

P = 0.99
Confidence = 0.5

N		δ = 5%	
1		1	
10		2	
100		2	
1,000		2	
1,000,000		2	
1,000,000,000		2	

Not that confident to start with

P = 0.99
Confidence = 0.5

N	δ = 10%	δ = 5%	δ = 1%
1	1	1	1
10	0	2	8
100	0	2	31
1,000	0	2	43
1,000,000	0	2	45
1,000,000,000	0	2	45

High confidence in statement

P = 0.5
Confidence = 0.99

N		δ = 5%	
1		1	
10		10	
100		87	
1,000		399	
1,000,000		663	
1,000,000,000		663	

High confidence in statement

$P = 0.5$
Confidence = 0.99

N	δ = 10%	δ = 5%	δ = 1%
1	1	1	1
10	9	10	10
100	63	87	99
1,000	142	399	943
1,000,000	166	663	16,317
1,000,000,000	166	663	16,587

End of second fast road...

Ommitted several details

- Mathematical model had several assumptions
- If your validation data doesn’t represent the population, nothing of what was said makes sense (the Math still works!)
- Please seek to understand more about causality, correlation, significance, random sampling, stratified sampling, cluster sampling, etc.

Testing ML requires domain knowledge

Make sure test data represents population

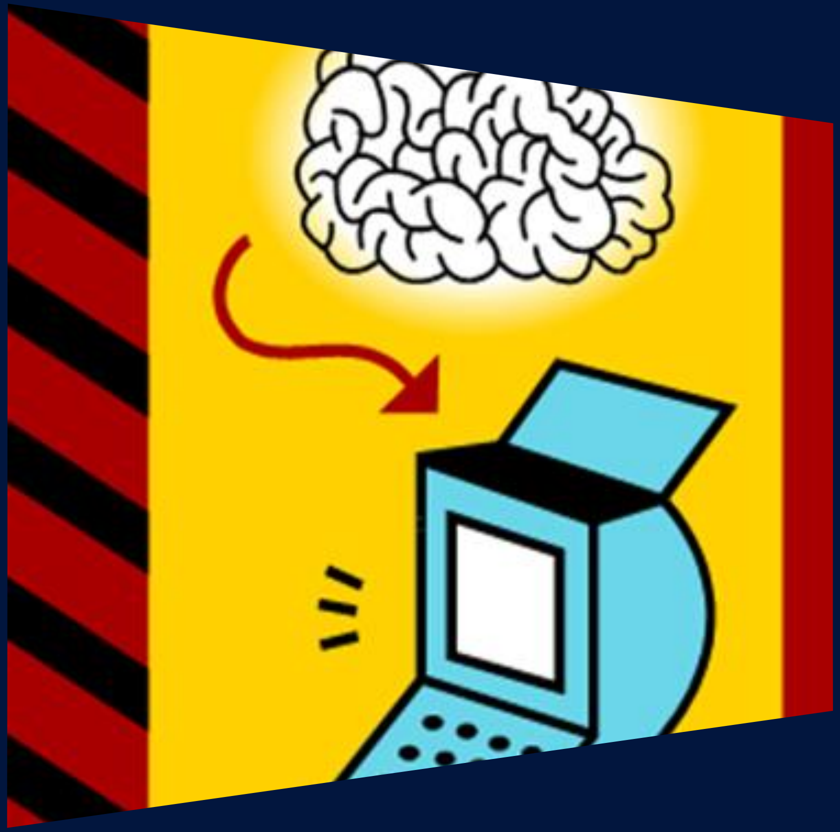
"Clustering" population

Characterizing classes of bugs properly

Otherwise you have bugs in quantity, but not in quality

LivePreview demo

Key messages



- Data isn’t the same as information
- Testing ML requires domain knowledge

Call to action and advice from experience

Browse through ML training online

Implement a heuristic solution before using ML

Develop plan to test your tools and solution

Microsoft