CLSAC
Chesapeake Large-Scale Analytics Conference

October 17-19, 2017

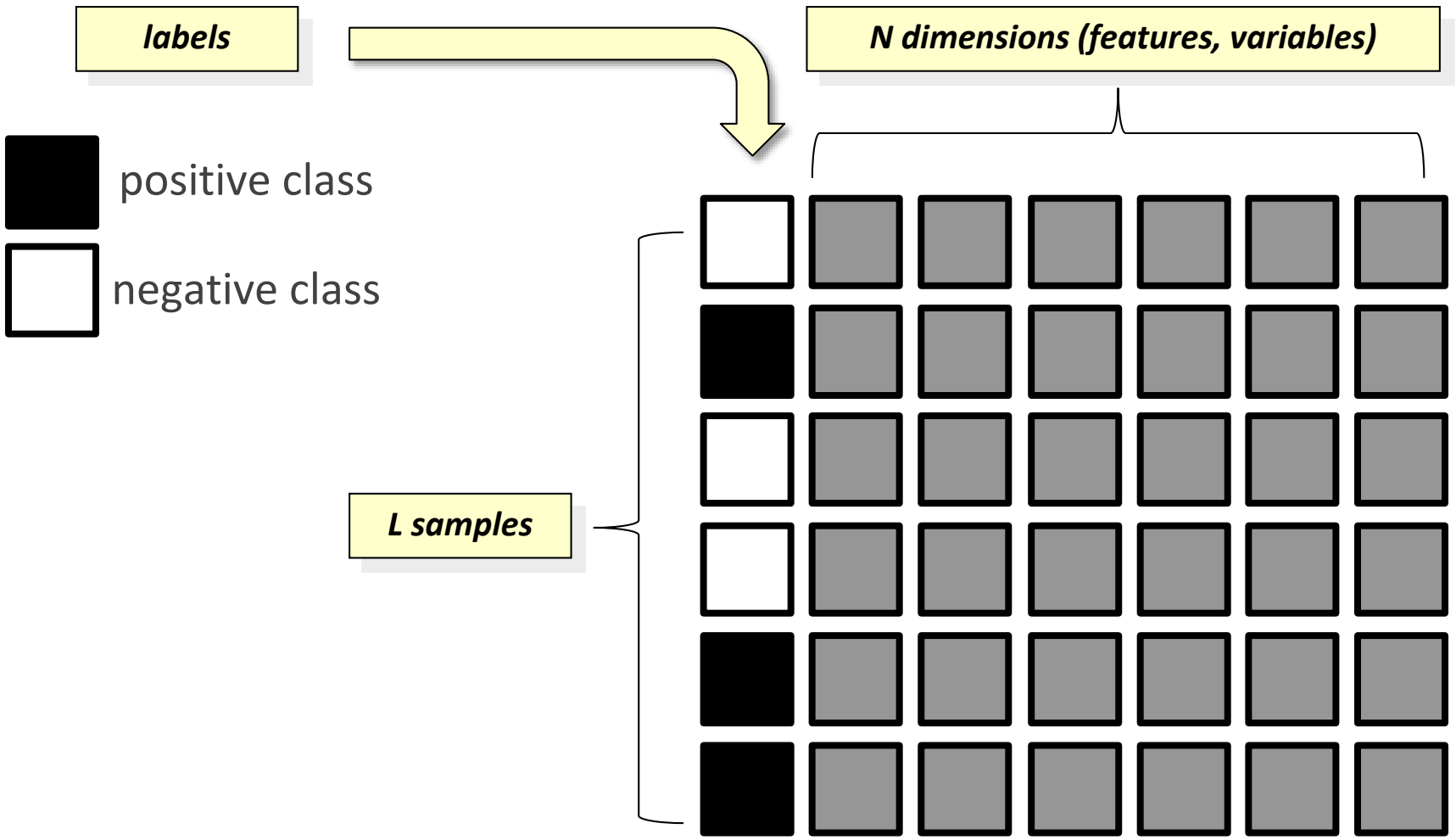
2017 Chesapeake Large-Scale Analytics Conference

Learning Using Privileged Information

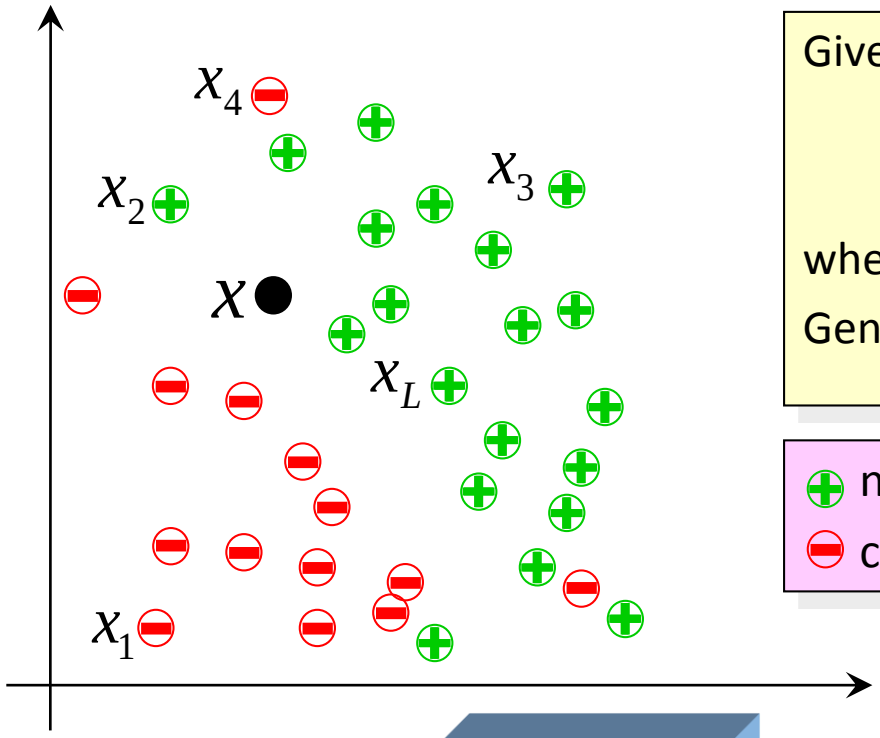
Rauf Izmailov
rizmailov@vencorelabs.com
Phone: 908-748-2891

VENCORELABS

Basic Binary Classification: Notations



Basic Binary Classification: Problem Formulation



Given training data (observations, facts)

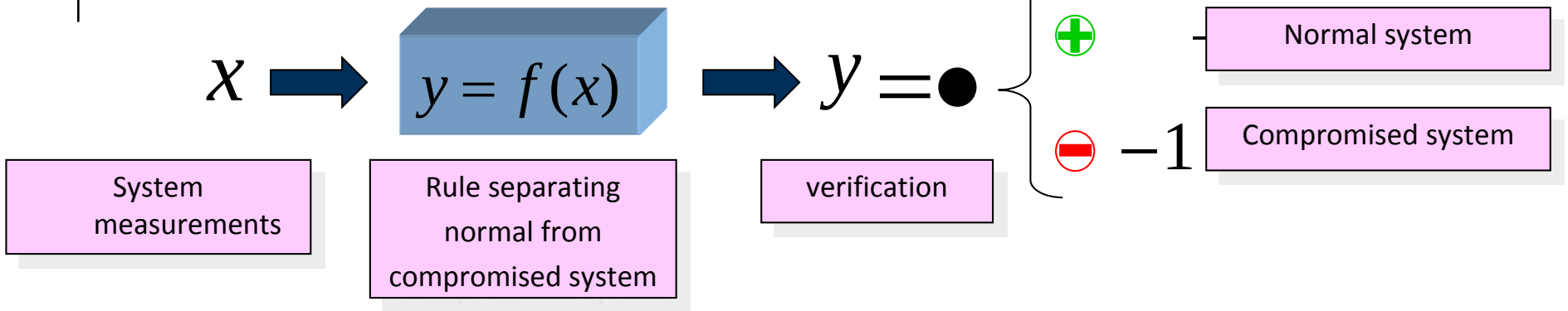
$$(x_1, y_1), \ldots, (x_L, y_L)$$

where $x \in R^n$ and $y \in \{-1, +1\}$

Generalize data to a rule (function $y = f(x)$

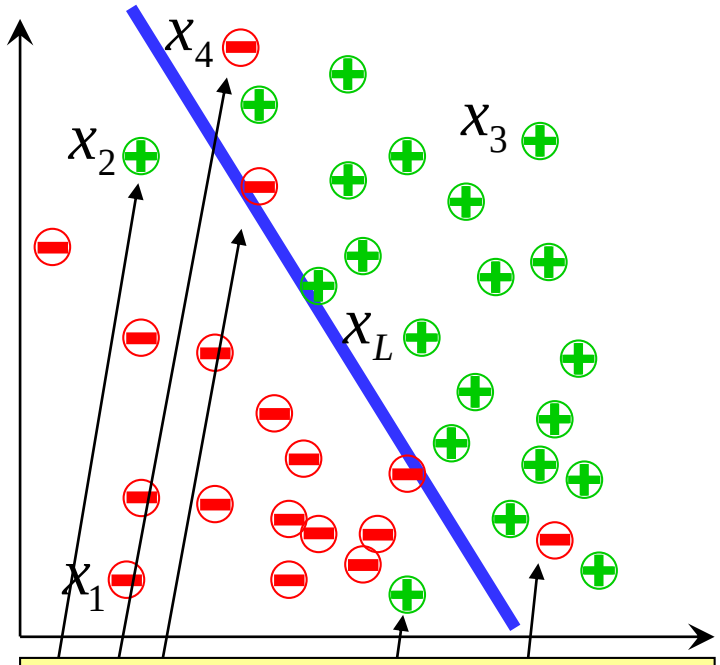
⊕ normal system

⊖ compromised system

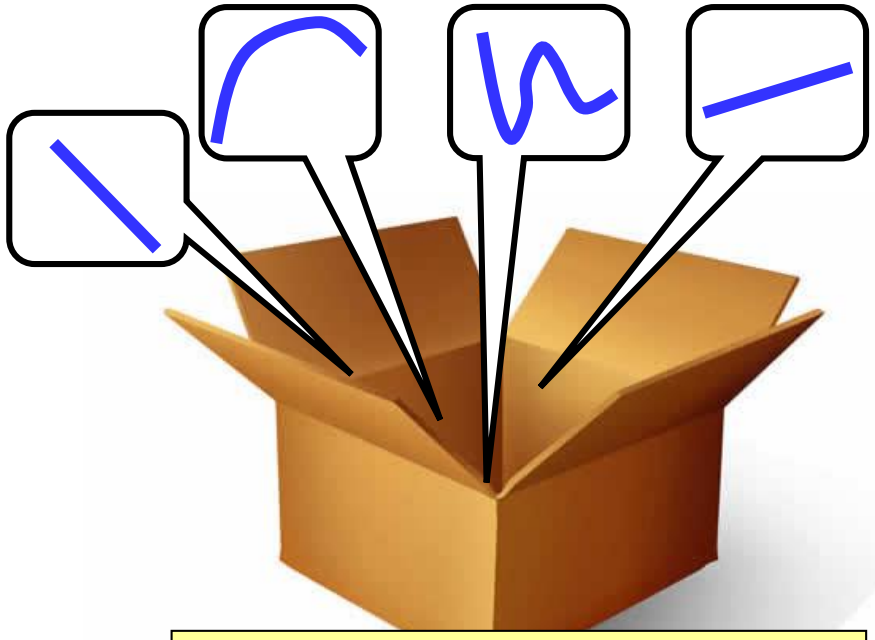


VENCORE LABS

Main Result of Classical Machine Learning Paradigm



errors on training sample



capacity of set of rules

There are **two** and **only two** factors responsible for induction:

1) Percentage of errors on the training sample (trivial factor)
2) Capacity (VC dimension) of the set of rules (non-trivial factor).

$$R(f) \le R_{\text{emp}}(f) + O\left(\sqrt{\frac{h}{L}}\right)$$

Confidence term:

- Increases with capacity (VC dimension) h
- Decreases with sample size L

Test error

Training error

VENCORELABS

Learning Problem

- Modern machine learning techniques for data analysis problems require construction of decision rules that operate in high dimensional spaces
  - To obtain good decision rules, one has to train learning algorithms using a huge number of data points
  - Meanwhile, humans can learn from a significantly smaller number of training examples
- Why the discrepancy?

Humans use a ***fundamentally different learning paradigm*** than machines

What's the difference?

Machine Learning Paradigm

Here are some examples of cats

Here are animals that are not cats

Learn a ***decision rule***:

INPUT → Decision Rule → OUTPUT

Not a cat

Human Learning Paradigm

Here are some examples of cats

Here are animals that are not cats

Some additional information about cats:

- Cute
- Tail
- Whiskers

ADDITIONAL (PRIVILEGED) INFORMATION

VENCORE LABS

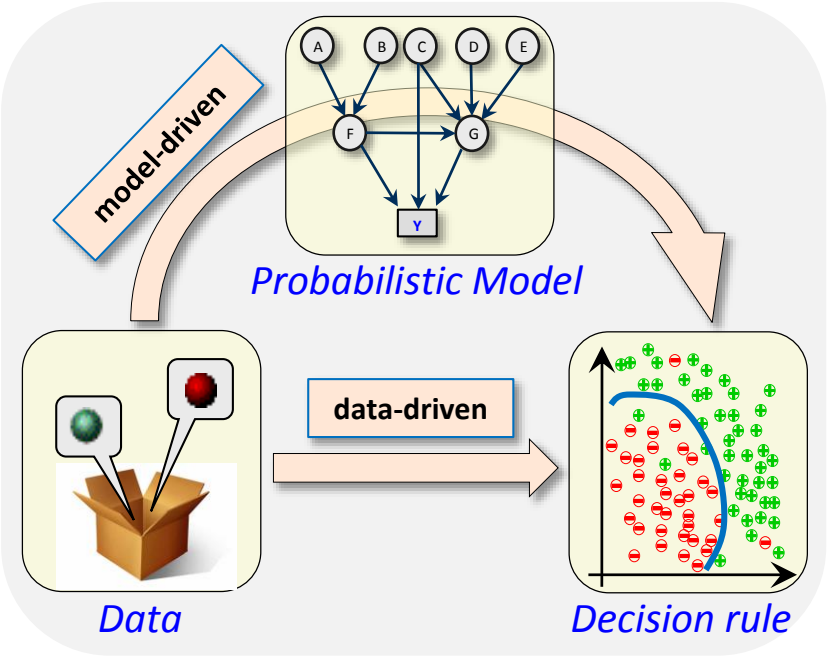
Model-Driven and Data-Driven Approaches

Problem of Model-driven approach:

- **Statistical method:** model is specified → estimate parameters → construct decision
- ***Advantages:***
  - Well-established mathematical tools to convert models into decision rules.
- ***Disadvantages:***
  - Actual system structure & distributions may differ from simplified model
  - Parameters may be hard to estimate (non-convex problem, insufficient data)
- ***Problem:*** *How to leverage* ***data-driven*** *information in* ***model-driven*** *setting?*

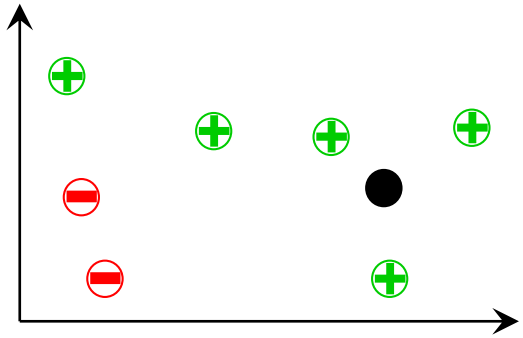
Problem of Data-driven approach:

- **Optimization method:** model is not specified => find the best approximation function
- **Advantages:**
  - Direct method (**one** hop versus **two** hops in model-driven approach)
  - Typically better performance than model-driven approach
- **Disadvantages:**
  - Ignores domain knowledge information on real structures
- ***Problem:*** *How to leverage* ***model-driven*** *information in* ***data-driven*** *approach?*



VENCORELABS

Learning Using Privileged Information (LUPI)



- Given training data (observations, facts)

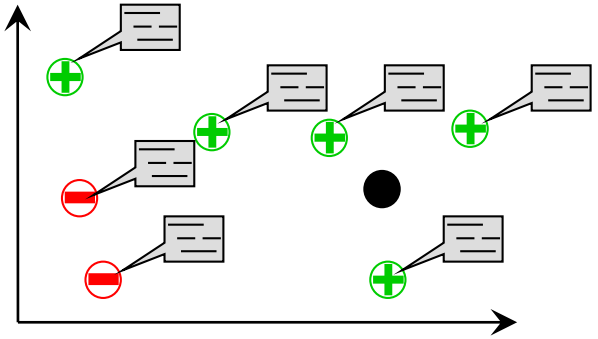
$$(x_1, y_1),\ldots,(x_L, y_L)$$

- Generalize data to a rule (function)

$$y = f(x)$$

- where $x \in X$ and $y \in \{-1,+1\}$

- Classical pattern recognition problem: training data and test data are from the same space, with have same attributes etc.



- Given training data (observations, facts)

$$(x_1, y_1),\ldots,(x_L, y_L)$$

and additional privileged data

$$x_1^*,\ldots x_L^*$$

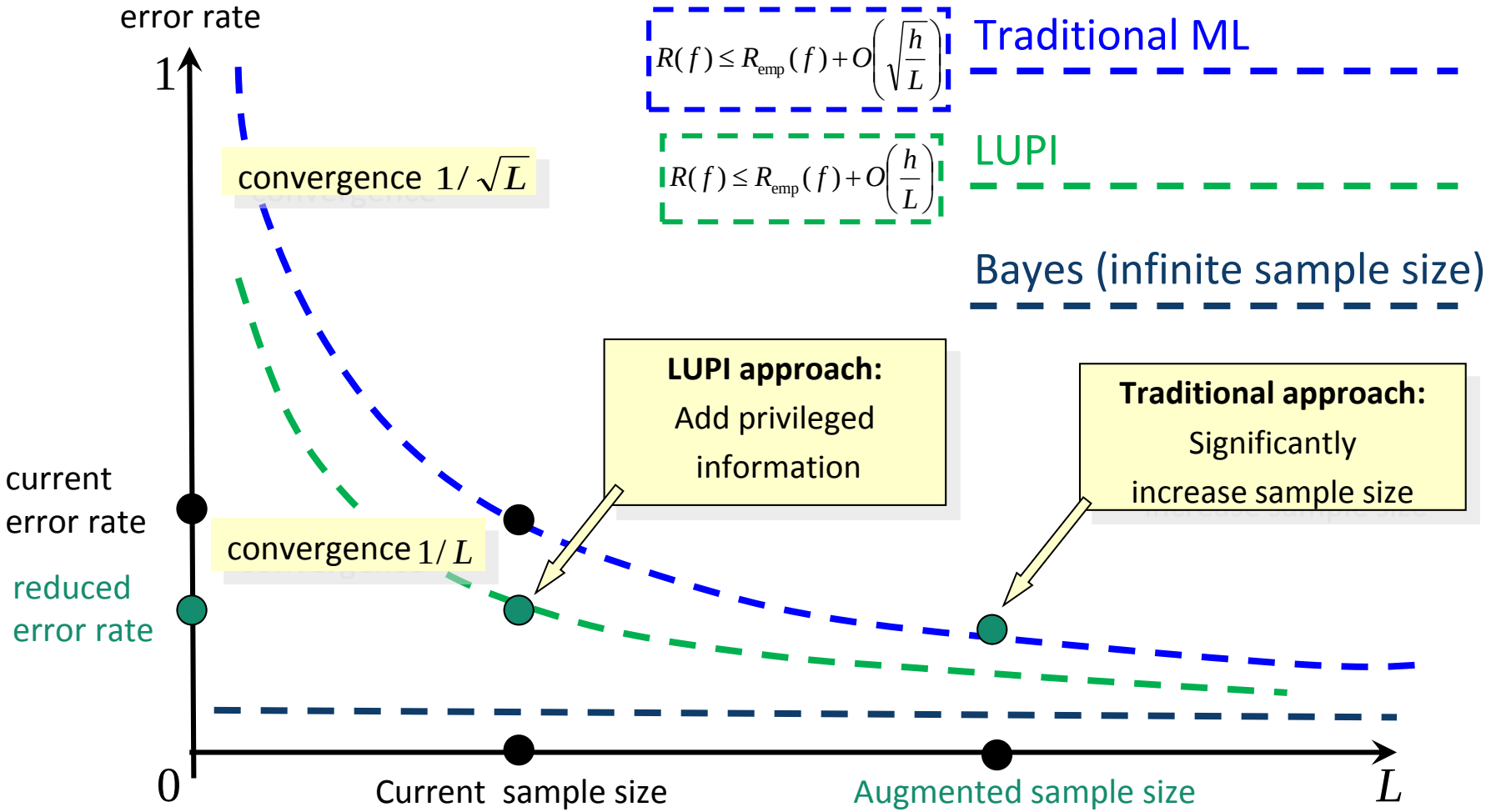
- Generalize data to a rule (function)

$$y = f(x)$$

- where $x \in X, x^* \in X^*$ and $y \in \{-1,+1\}$

- New paradigm of learning with privileged information: additional information is available **ONLY** with training data, but NOT with test data

Why Bother With LUPI?



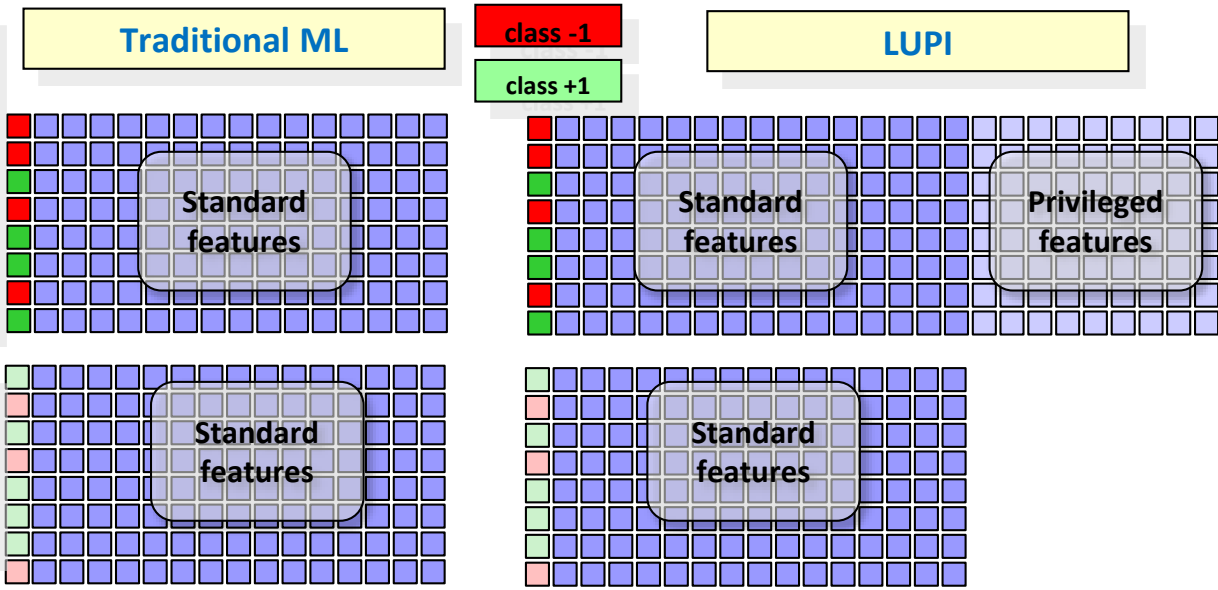
- SVM converges (as $1/\sqrt{L}$) to the default Bayes rule as sample size L goes to infinity
- In reality, almost all samples are small; additional samples are expensive/impossible to get
- LUPI uses privileged info instead of additional samples to converge much faster (as $1/L$)

VENCORELABS

Learning Using Privileged Information (LUPI)

Training data:
- Off-line processing
- High-quality data
- Additional features used as privileged information

Test data:
- On-line processing
- Reduced-quality data



standard features

privileged features

LUPI uses fundamental asymmetry between training & test data and leverages high-quality privileged information available during training for better performance

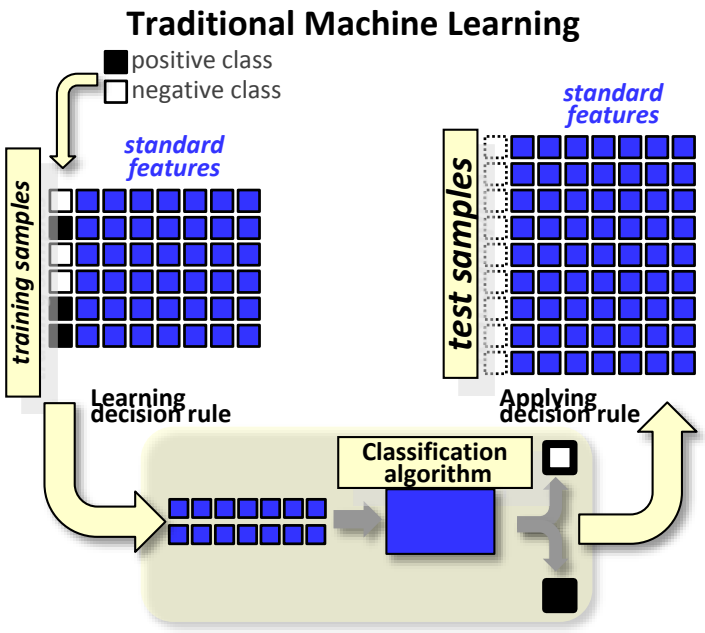
LUPI converges to the solution much faster than alternatives (needs 33 examples instead of 1,000, or 100 instead of 10,000)

Privileged data have the same properties as labels

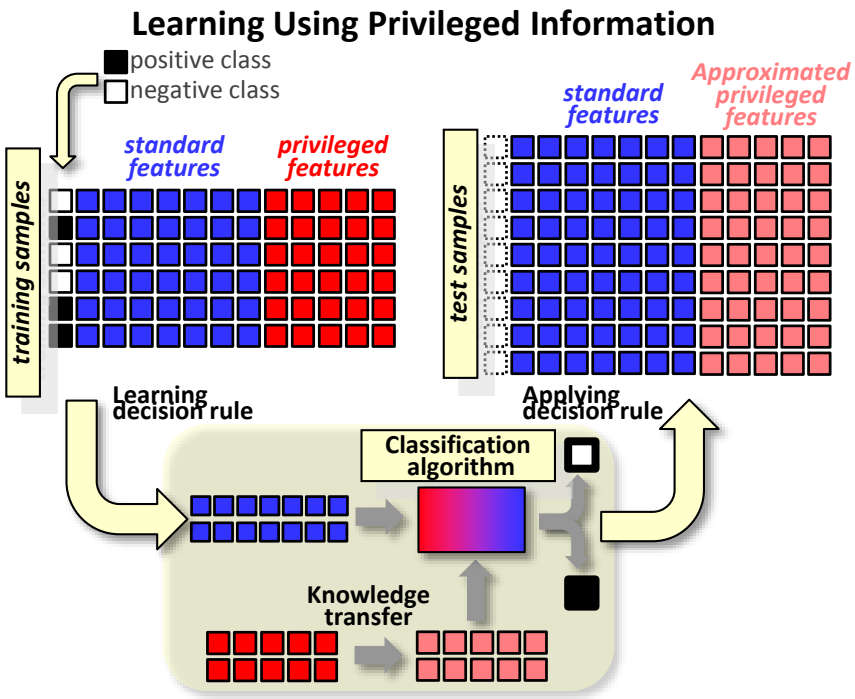
Standard: hand-written digits and their classification
Privileged: Semantic description of digits shapes
Output: classification of pixel image into 0,1,2,…,9
Decision rule: classification based **ONLY** on pixel image

VENCORE LABS

General LUPI Mechanism



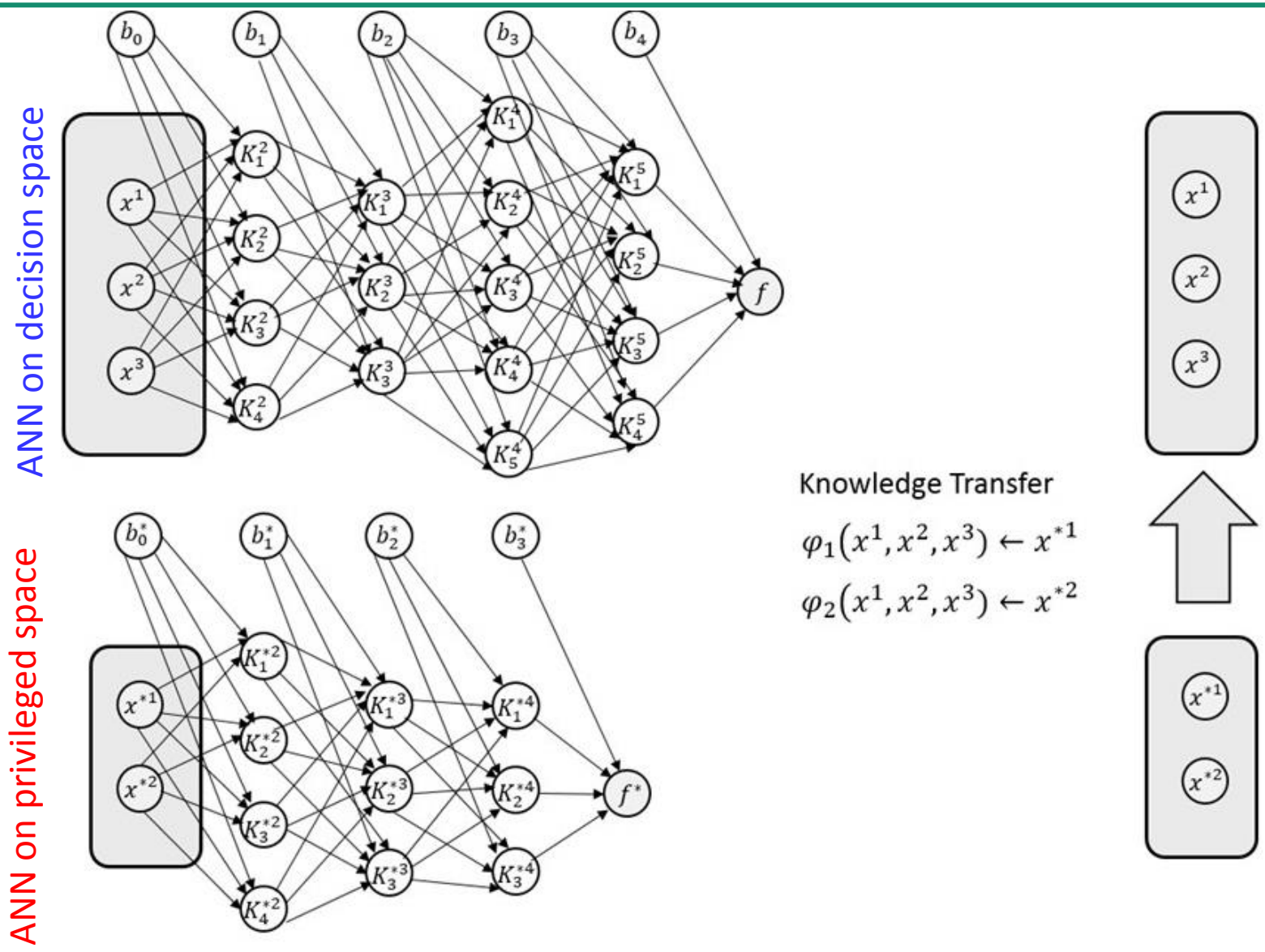
- Rule is learned only on **standard** features
- Rule works only on **standard** features
- **Privileged** features, if they exist, are ignored



- Rule learned on **both** **standard** and **privileged** features
- **Privileged** features **partially** learned (approximated) from **standard** ones
- Works on **standard** features and approximated **privileged** ones

VENCORELABS

LUPI Mechanism for Neural Networks



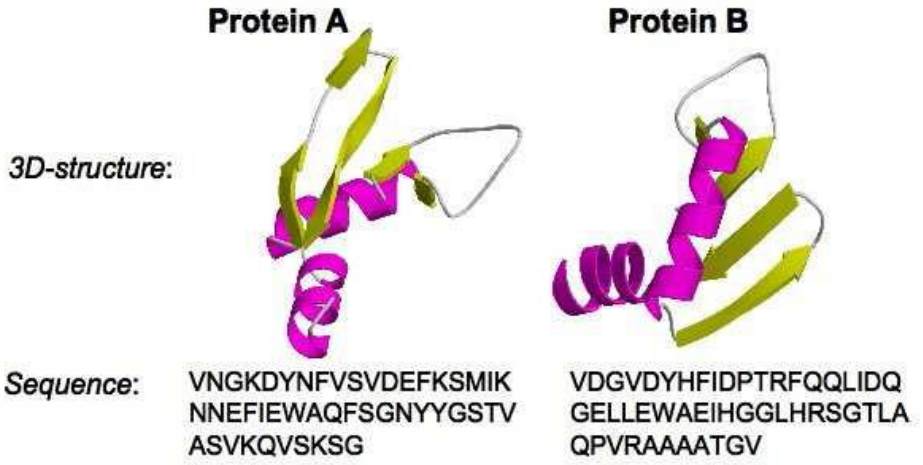
VENCORELABS

Some Examples of LUPI Applications

Privileged information	*Examples*
Future events	*LUPI has been applied to prediction of quasi-chaotic time series (future-in-the-past was privileged information)*
Detailed description of events (semantic information) produced by human experts	*LUPI has been applied to image classification (semantic description of images was privileged information)*
Time-consuming probing of data	*LUPI has been applied to protein classification (3D protein folding was privileged information)*
Heterogeneous sources of information, some of which may unavailable during test	*LUPI has been applied to human detection on a combination of electro-optical and infra-red sensors (one type of sensor was privileged)*
Expensive sensors	*LUPI has been applied to human detection on a combination of expensive (high quality) and checp (low quality) video cameras*

Advanced Model as Privileged Information

- Problem: given amino-acid sequences of proteins; classify them into families of proteins
- Assumptions:
  - The training data: the space of amino-acid sequences (relatively easy to obtain)
  - The privileged information: the space of 3D structures of the proteins (difficult to obtain)
- Source:
  - SCOP database (structural classification of proteins): sequences and their hierarchical organizations
  - 80 superfamilies with the largest number of sequences in each



3D-structures are close

sequences may not be close

Classification or Protein Families

Protein superfamily pair	SVM	LUPI	SVM (full)	
a.26.1-vs-c.68.1	7.3	7.3	0	
a.26.1-vs-g.17.1	16.4	14.3	0	
a.118.1-vs-b.82.1	19.2	6.4	0	★
a.118.1-vs-d.2.1	41.5	24.5	3.8	
a.118.1-vs-d.14.1	13.1	13.1	2.2	
a.118.1-vs-e.8.1	22.8	2.3	2.3	★
b.1.18-vs-b.55.1	14.6	13.5	0	
b.18.1-vs-b.55.1	31.5	15.1	0	★
b.18.1-vs-c.55.1	36.2	36.2	0	★
b.18.1-vs-c.55.3	38.1	36.6	0	★
b.18.1-vs-d.92.1	25	11.8	0	★
b.29.1-vs-b.30.5	16.9	16.9	3.6	
b.29.1-vs-b.55.1	10	5.5	0	
b.29.1-vs-b.80.1	8.3	5.9	0	
b.29.1-vs-b.121.4	35.9	16.8	5.3	★

Protein superfamily pair	SVM	LUPI	SVM (full)	
b.29.1-vs-c.47.1	10.2	3.2	0	★
b.30.5-vs-b.80.1	43.3	6.7	0	★
b.30.5-vs-b.55.1	25.5	14.6	0	
b.55.1-vs-b.82.1	11.8	10.3	0	
b.55.1-vs-d.14.1	20.9	19.4	0	
b.55.1-vs-d.15.1	17.7	12.7	0	
b.80.1-vs-b.82.1	4.7	4.7	0	
b.82.1-vs-b.121.4	7.9	3.4	0	★
b.121.4-vs-d.14.1	29.5	23.9	0	
b.121.4-vs-d.92.1	15.3	9.2	0	
c.36.1-vs-c.68.1	8.9	0	0	★
c.36.1-vs-e.8.1	12.8	2.2	0	★
c.47.1-vs-c.69.1	1.9	0.6	0	★
c.52.1-vs-b.80.1	11.8	5.9	0	
c.55.1-vs-c.55.3	45.1	28.2	22.5	

★ 3D structure is essential for classification; SVM+ does not improve classification of SVM

★ SVM+ provides significant improve improvement over SVM (several times)

VENCORE LABS

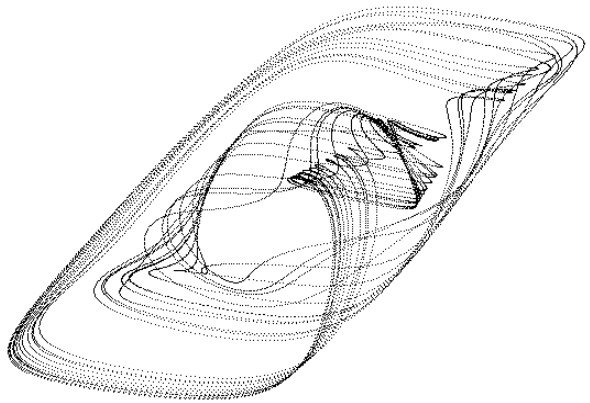
Future Events as Privileged Information

- The goal is to predict the development of the system (time series) at the specified time in the future
- Privileged data include the evolution / trajectory of the system from current moment to the targeted time in the future
- Given archived data, privileged data can be viewed as “future-in-the-past”

Data generated by the Mackey-Glass equation:

$$\frac{dx(t)}{dt} = -ax(t) + \frac{bx(t-\tau)}{1+x^{10}(t-\tau)},$$

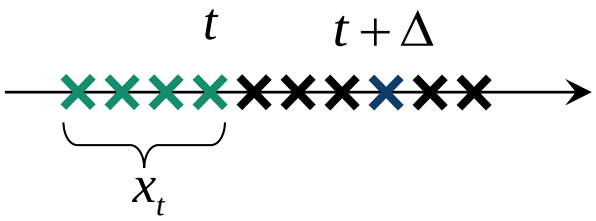
where $a, b,$ and τ (delay) are parameters.



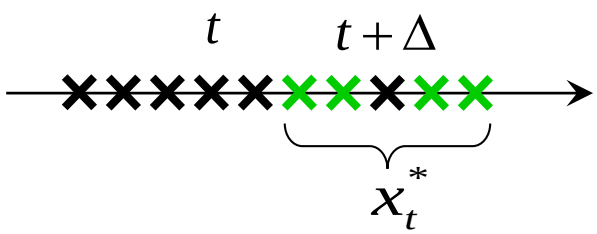
The training triplets:

$$x_t = (x(t), x(t-1), x(t-2), x(t-3))$$

$$x_t^* = (x(t+\Delta-1), x(t+\Delta-2), x(t+\Delta+1), x(t+\Delta+2))$$



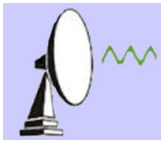
current value and past values



values in future (around Δ)

VENCORE LABS

LUPI for Decision Making with Unreliable Sensors

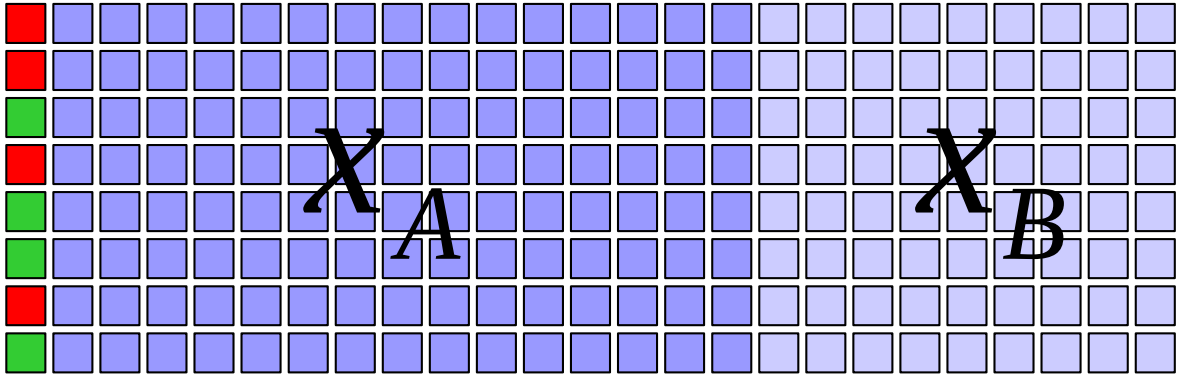


video (A)

radar (B)

class -1

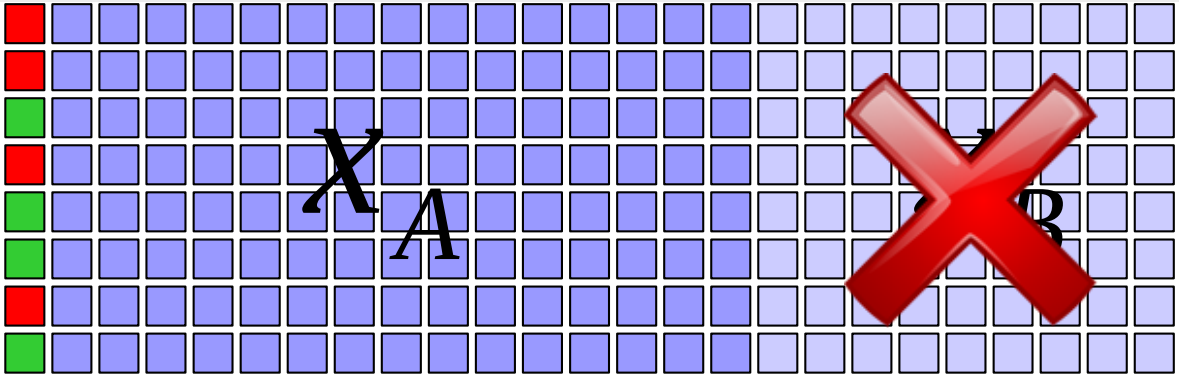
class +1



If one of the sensors is down, the corresponding features are not available.

class -1

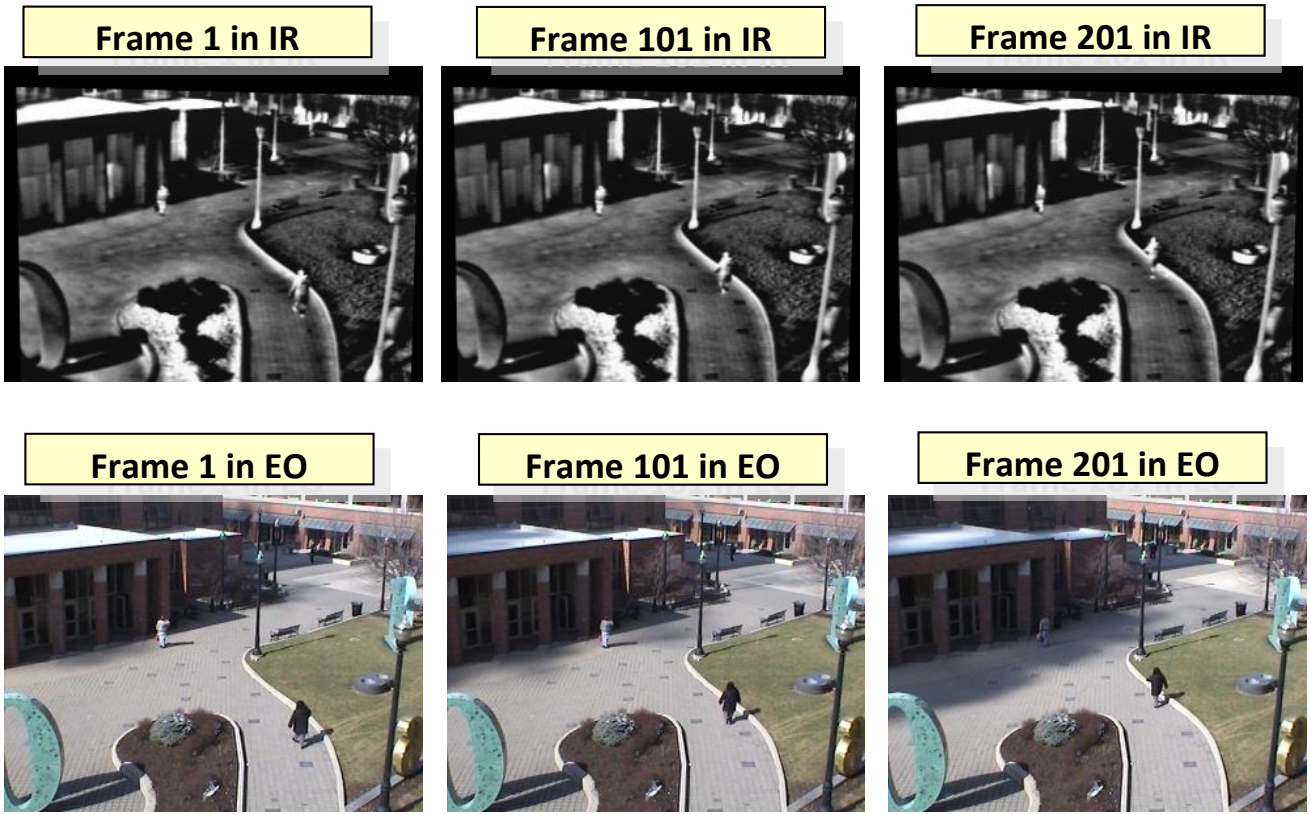
class +1



VENCORELABS

Example: EO/IR Monitoring

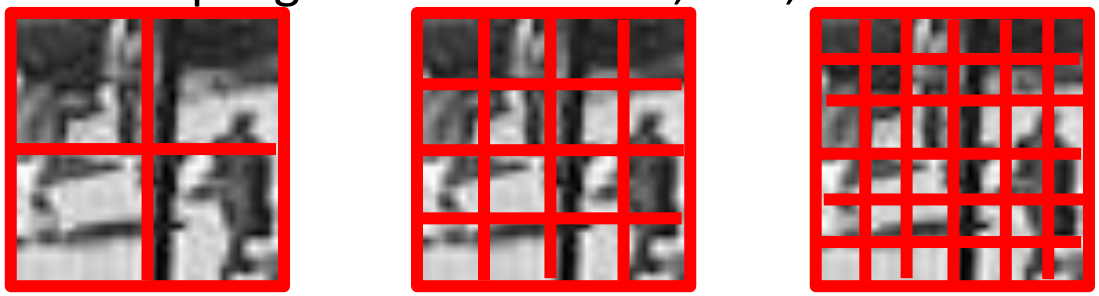
- EO/IR benchmark dataset from OSU
- 3 paired (EO and IR) surveillance videos in the form of sequential images (total number of frames about 8,000)
- The goal is to detect humans on video



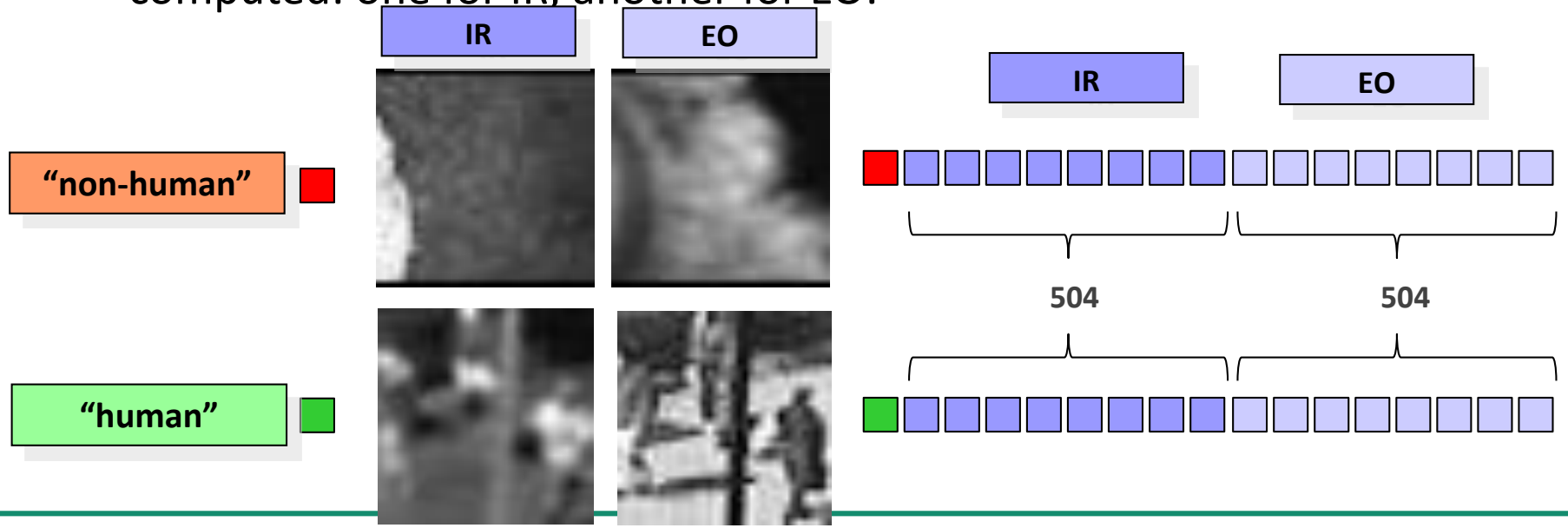
VENCORELABS

Feature Generation

- Each 36*36 image was converted to a string of HOG features using partitions of multiple granularities: 2*2, 4*4, and 6*6:

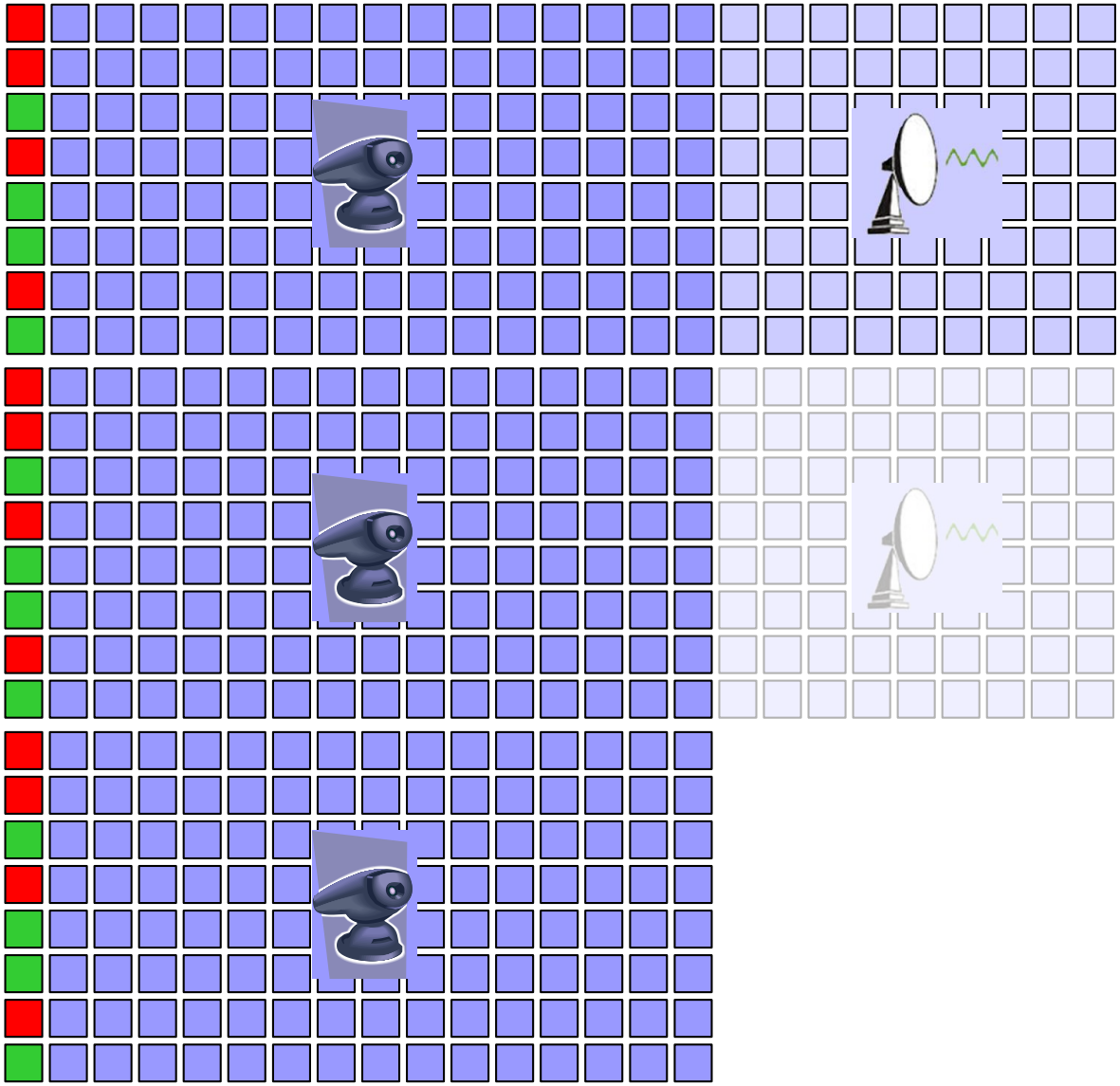


- For each image, two 504-dimensional vectors of features were computed: one for IR, another for EO:



VENCORE LABS

LUPI Application for Missing Data



Current ML paradigm:
use both available sensors

error rate

LUPI paradigm:
use unavailable sensor

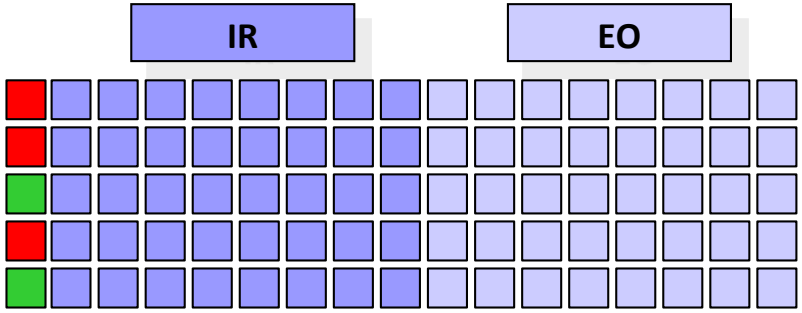
error rate

Current ML paradigm:
use one available sensor

error rate

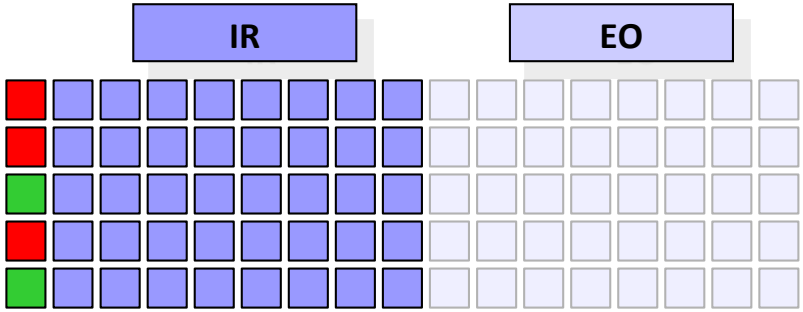
VENCORELABS

Classification Decision Results



Current ML paradigm:
use both available sensors

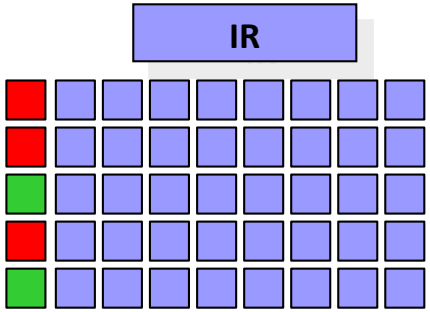
error rate
14.62%



error rate
reduction
by 33%

LUPI paradigm:
use unavailable sensor

error rate
16.10%

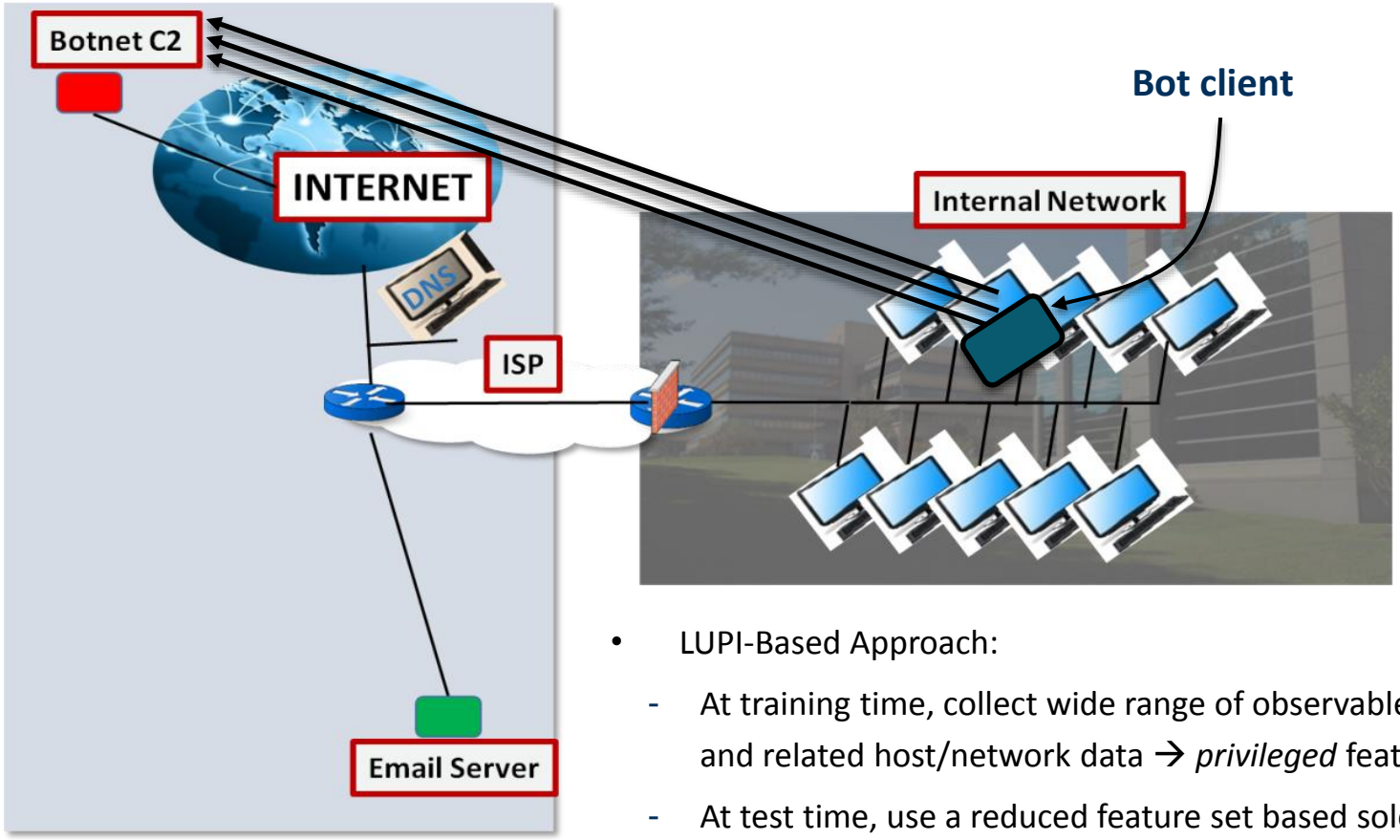


Current ML paradigm:
use one available sensor

error rate
16.85%

VENCORELABS

Application of LUPI to Cyber Analytics



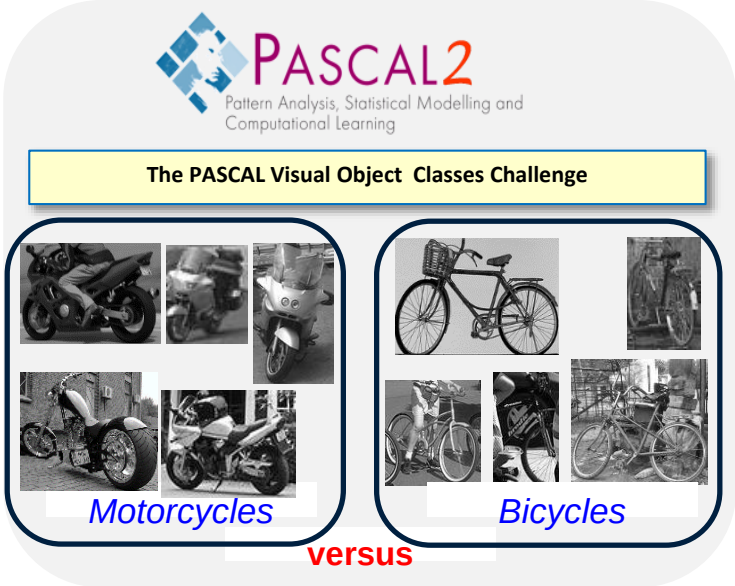
- LUPI-Based Approach:
  - At training time, collect wide range of observables, including user behavior and related host/network data → *privileged* features used only for training
  - At test time, use a reduced feature set based solely on traffic generated by host that is guaranteed to be observable from outside the host

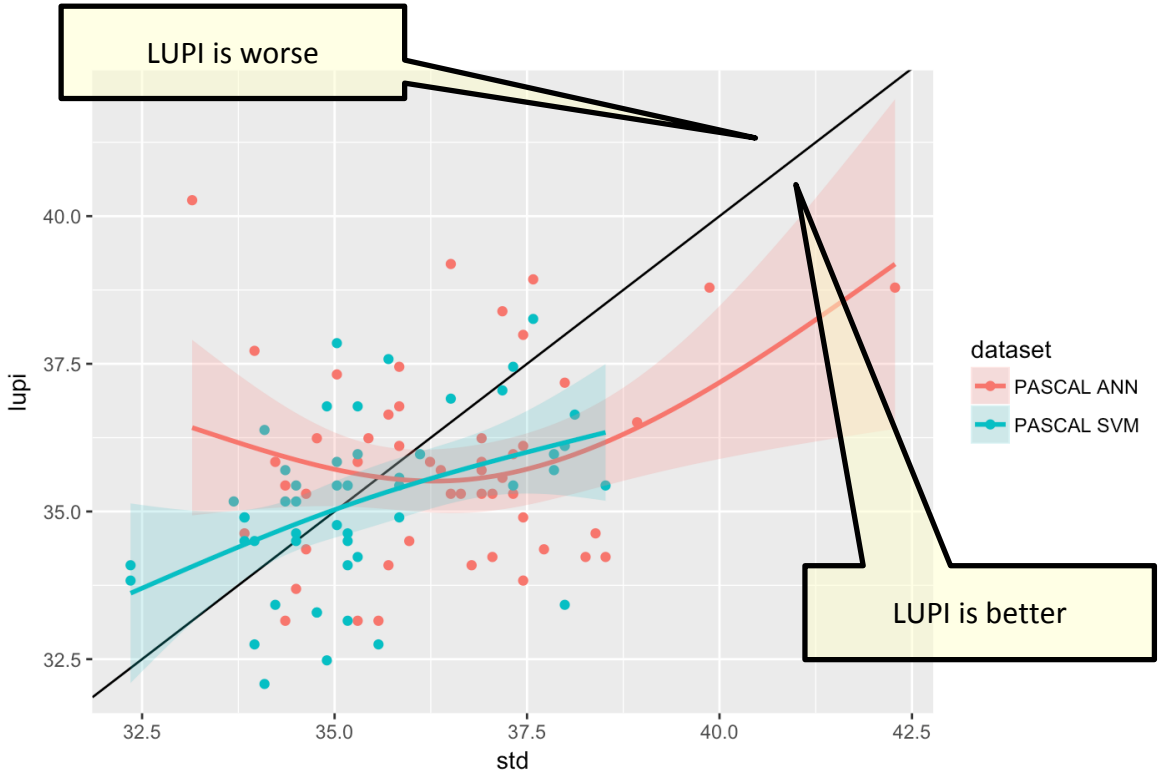
Results: Order of magnitude higher detection accuracy using LUPI

VENCORELABS

Application of LUPI to Image Classification

- Data: 20 classes of objects (cows, bicycles, dogs, cars, cats, etc.
- Objects downloaded, extracted and mapped to HOG features
- Selected two (arguably, most difficult in terms of separation) classes (motorcycles and bicycles)
- Effect of privileged features emulated by obscuring half of the image





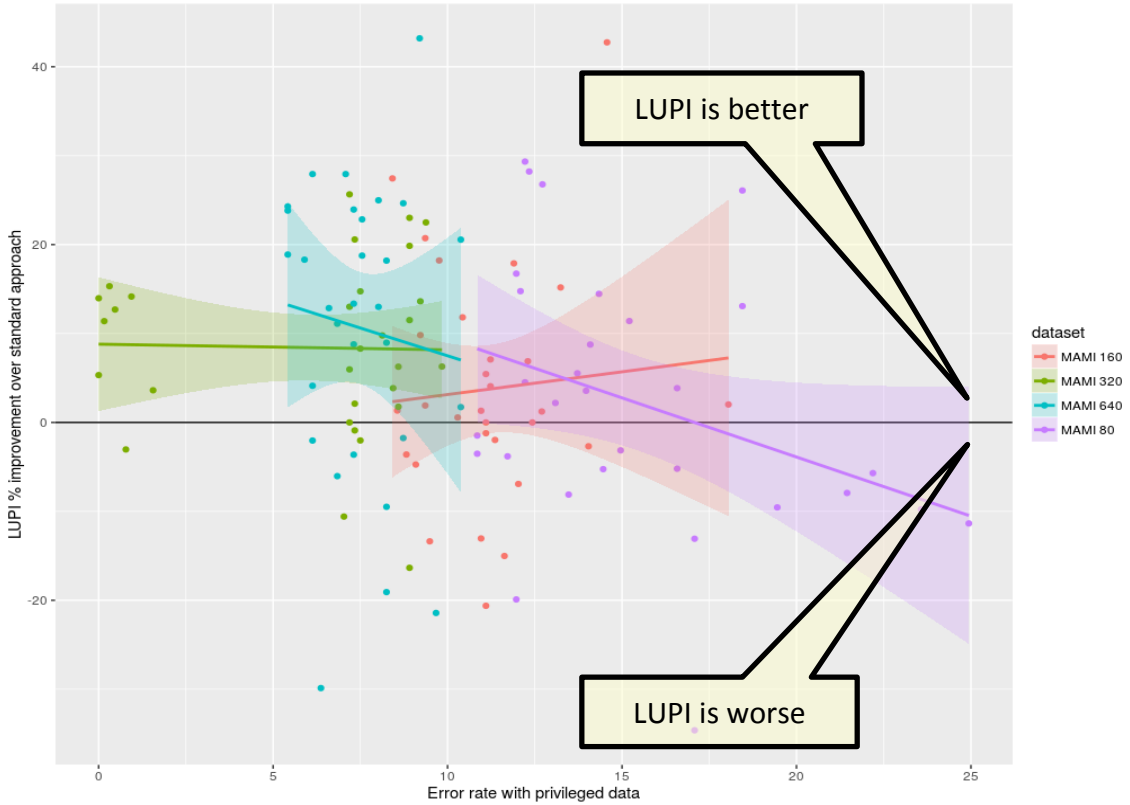
VENCORELABS

Application of LUPI to Target Recognition

- Minor Area Motion Imagery (MAMI) dataset (AFRL, 2014)
- Standard/Privileged features correspond to different resolution/quality of videos
- Consistent LUPI performance advantage for larger training sizes
- Partial LUPI performance advantage for small training size: 80
- *Exception: high error rate of standard case **and** small training size*

Standard

Privileged



VENCORELABS

LUPI: Summary

- LUPI was first introduced about 10 years ago. Initial LUPI framework (called SVM+) was limited just to SVM architecture, and had limited scalability up to 200-300 points in the training dataset (the corresponding matrix was ill-conditioned)
- Current LUPI framework is as scalable as standard classifiers, not restricted to SVM (LUPI works for neural networks, etc.)
- Wide scope of LUPI applications (a few calibration sets already made public).
- Reasonable performance gains of LUPI (20%-50%)
- Open Source version of LUPI (in scikit-learn) is being developed within a DARPA program and to be released soon.
- First Workshop on Privileged Information "Beyond Labeler" last year: http://smileclinic.alwaysdata.net/ijcai16workshop/
- Active on-going research in privileged information:
  - Mechanisms for approximation of privileged information
  - Mechanisms for selection/filtering of privileged information
  - Expansion to non-supervised learning applications

LUPI References

- D.Pechyony, R.Izmailov, A.Vashist, V.Vapnik, SMO-style Algorithms for Learning Using Privileged Information, in Proceedings of the 2010 International Conference on Data Mining (DMIN), 2010.
- V.Vapnik, R.Izmailov, Learning Using Privileged Information: Similarity Control and Knowledge Transfer, Journal of Machine Learning Research, 16:2023-2049, 2015.
- V.Vapnik, R.Izmailov, Learning with Intelligent Teacher: Similarity Control and Knowledge Transfer, in Statistical Learning and Data Sciences, A.Gammerman, V.Vovk, H.Papadopulos (Eds). Lecture Notes in Artificial Intelligence 9047. Proceedings of Third International Symposium, SLDS. London, Springer, 2015, pp.3-32.
- V.Vapnik, R.Izmailov, Learning with Intelligent Teacher, in Lecture Notes in Artificial Intelligence 9653. Proceedings of 5th International Symposium, COPA 2016. Springer, 2016.
- R.Ilin, S.Streltsov, R.Izmailov, Learning with Privileged Information for Improved Target Classification, International Journal of Monitoring and Surveillance Technologies Research, 2(3), 5-66, July 2014.
- R.Ilin, R.Izmailov, Y.Goncharov, S.Streltsov, Fusion of Privileged Features for Efficient Classifier Training, 19th International Conference on Information Fusion, pp.1-8, 2016.
- V.Vapnik, R.Izmailov, Knowledge Transfer in SVM and Neural Networks, Annals of Mathematics and Artificial Intelligence, 1-17, 2017.
- A.Sapello, C.Serban, R.Chadha, R.Izmailov, Application of Learning Using Privileged Information (LUPI): Botnet Detection, Proceedings of 26th International Conference on Computer Communication and Networks (ICCCN), 2017.
- Z. Celik, P. McDaniel, R. Izmailov, Feature Cultivation in Privileged Information-augmented Detection, Proceedings of IWSPA 2017, 3rd International Workshop on Security and Privacy Analytics, pp.73-80, 2017.
- R.Izmailov, B.Lindqvist, P.Lin, Feature Selection in Learning using Privileged Information, Proceedings of IEEE International Conference on Data Mining (ICDM) 2017.

This material is based upon work partially supported by AFRL under contract FA9550-15-1-0502. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

VENCORE LABS

TRANSFORMATIVE RESEARCH