Tutorial talk, May 14th 2018

POLITECNICO DI MILANO

An Introduction to Machine Learning in Optical Transport networks

Massimo Tornatore

Department of Electronics, Information and Bioengineering
Politecnico di Milano, Italy

Covered topics

- The presentation is organized into two main parts
- Part 1: overview on Machine Learning
  - Basic concepts (supervised/unsupervised learning, bias/variance trade-off, etc.)
  - Some algorithms
    - Linear regression
    - Neural Networks
    - K-nearest neighbours

Note: this is NOT a "pure" Machine Learning tutorial

The objective is to show **how we applied** ML to **our** research problems

- Part 2: applications of ML to optical-network problems
  - Part 2a): QoT estimation and RSA
  - Part 2b): Failure management
  - Part 2c): Other application at physical and networklayer
    - Traffic prediction, virtual topology design,…

What is Machine Learning?

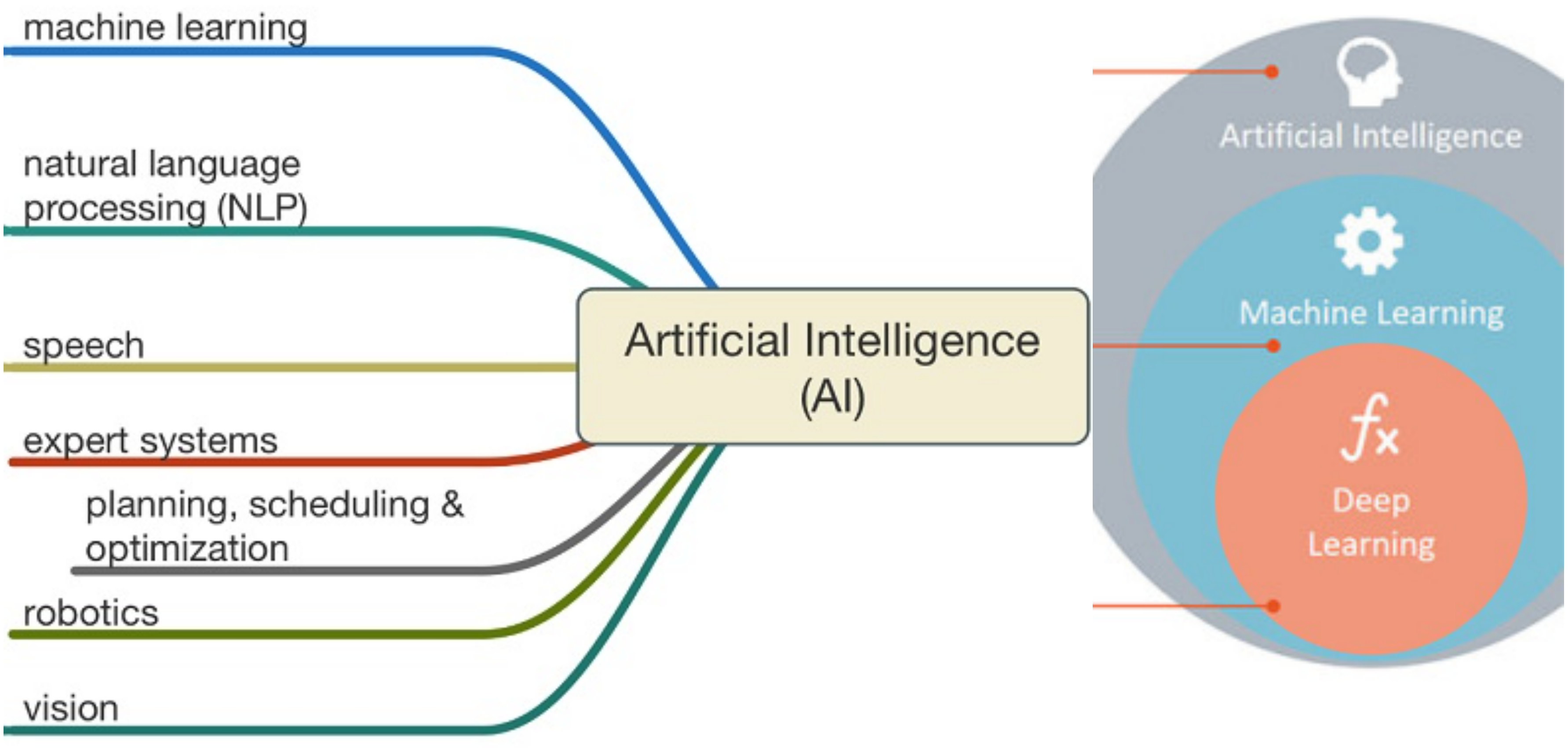
- *"Field of study that gives computers the ability to learn without being explicitly programmed" (A. Samuel, 1959)*
- *"Teaching a computer to automatically learn concepts through data observation"*
- *…*

- For our purposes: An math/statistical instrument to make decisions by inferring statistical properties of monitored data

  …in the context of optical networks

- Sometimes confused with other terms: AI, Deep Learning, Data Analytics, Data Mining, etc.

Many definitions with blurred borders



Why only now in optical networks?

- At data plane - Complexity increase
  - Coherent Trasmission System
    - Several system parameters to choose from: modulation techniques and formats, coding rates, symbol rate..
    - DSP: Huge availability of data
  - Elastic Optical Networks
    - Customizable channel width, BV-ROADM
- At control plane - New Enablers
  - Software Defined Networking
  - Intelligence (computing capabiltiies) everywhere (e.g., MEC)
  - Monitors

Main categories of ML algorithms (1)

- Supervised-learning algorithms
  - We are given “labeled” data (i.e., “ground truth”)
  - Main objective: given a set of “historical” input(s) predict an output
    - Regression: output value is continuous
    - Classification: output value is discrete or “categorical”

- An example: Traffic forecasts
  - Given traffic during last week/month/year
    - Predict traffic for the next period (regression)
    - Predict if available resources will be sufficient (classification)

- Other examples
  - Speech/image recognition
  - Spam classifier
  - House prices prediction/estimation

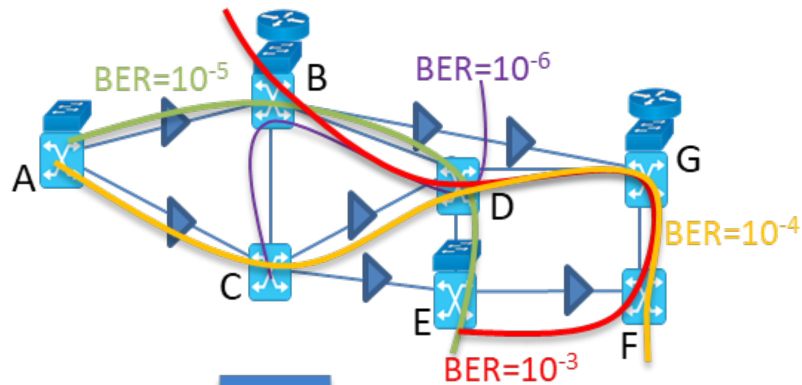
Supervised learning: an «optical» example

Training Phase

λ=1550, path= nodes A-C-D-G-F,
Mod = QPSK, ➔ BER=10^{-5}

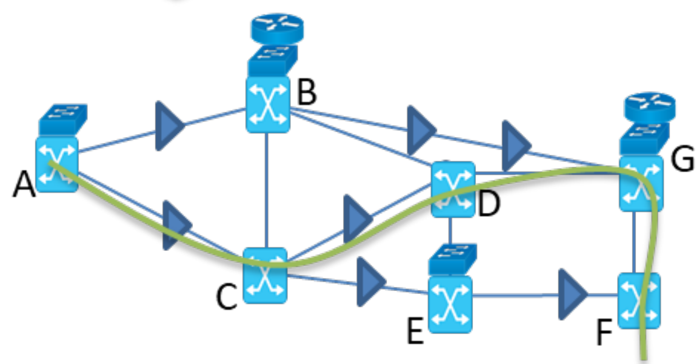
λ=1553, path= nodes B-G-D-F-E,
Mod = QPSK, ➔ BER=10^{-2}

...



Active/Test Phase

Create path: λ= 1553, nodes A-C-D-G-F,
Mod QPSK ➔ BER=?



Courtesy of Marco Ruffini and Irene Macaluso

Supervised Learning: the algorithm is trained on dataset that consists of paths, wavelengths, modulation, and the corresponding BER. Then it extrapolates the BER in correspondence to new inputs.

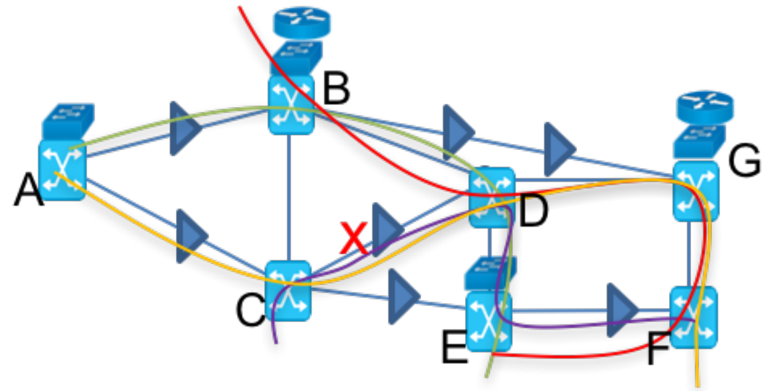
Main categories of ML algorithms (2)

- Unsupervised-learning algorithms
  - Available data is not "labeled"
  - Main objective: derive structures (patterns) from available data
    - Clustering finding "groups" of similar data
    - Anomaly detection

- An example: cell-t+-raffic classification
  - Given traffic traces i
  - understand if some cells provide similar patterns
    - Residential, business, close to theatre, cinema, stadium…
    - This information can be used to make network resources planning

- Other example
  - Group people according to their interests to improve advertisement

Unsupervised learning: some examples



Data:

λ=1550, path= nodes A-B-D-E, Mod = QPSK, BER=10^{-6}

λ=1553, path= nodes A-C-D-G-F, Mod = BPSK, BER=10^{-2}

λ=1544, path= nodes C-D-E-F, Mod = DPQPSK, BER=10^{-2}

λ=1545, path= nodes B-D-G-F-E, Mod = 16-QAM, BER=10^{-7}

Anomaly detection for link C-D

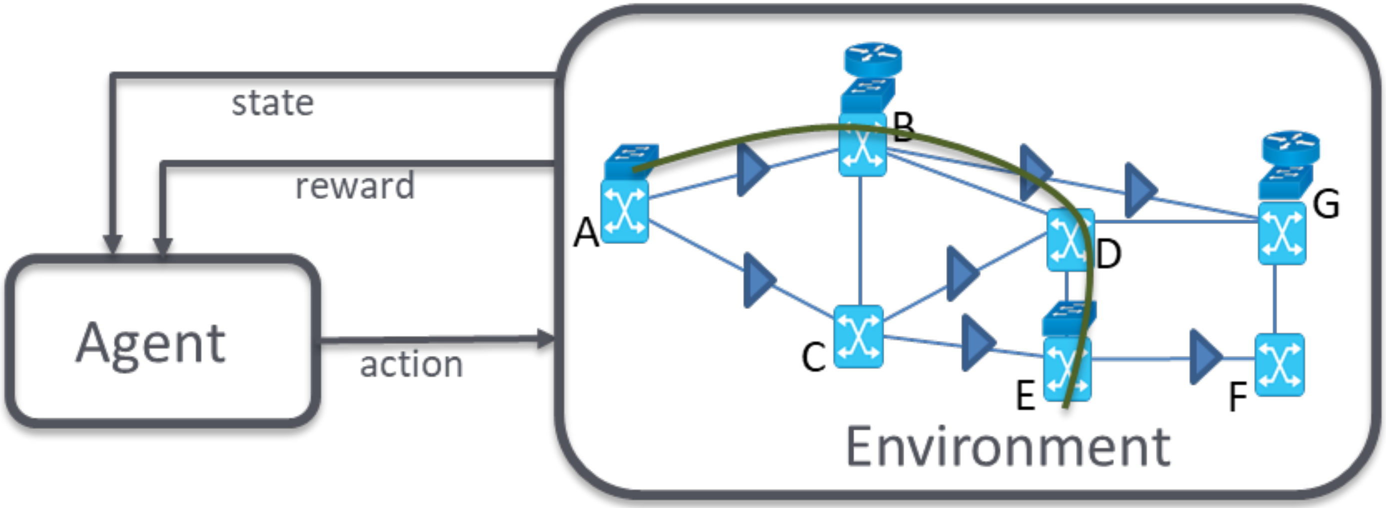
Courtesy of Marco Ruffini and Irene Macaluso

Unsupervised Learning: the algorithm identifies unusual patterns in the data, consisting of wavelengths, paths, BER, and modulation..

Main categories of ML algorithms
Cont'd

- Semi-Supervised learning
  - Hybrid of previous two categories
  - Main objective: most of the training samples are unlabeled, only few are labeled.
    - Common when labeled data are scarce or expensive
  - Self-training: start with labeled data, then label unlabeled data based on first phase
- Reinforcement learning
  - Available data is not "labeled"
  - Main objective: learn a policy, i.e., a mapping between in inputs/states and actions. Behavior is refined through rewards
  - Methodologically similar to «optimal control theory» or «dynamic programming»
  - Q-learning

Reinforcement learning: some examples



Initial state	Action	State	Reward
λ=1550nm, nodes A-B-D-E, Mod QPSK, BER=10^{-3}	Change: Mod DPQPSK	BER= 10^{-3}	0
λ=1550nm, nodes A-B-D-E, Mod QPSK, BER=10^{-3}	Change: output power channel +5 dBm	BER= 10^{-2}	-1
λ=1550nm, nodes A-B-D-E, Mod QPSK, BER=10^{-3}	Change: Mod BPSK	BER= 10^{-4}	+1

Courtesy of Marco Ruffini and Irene Macaluso

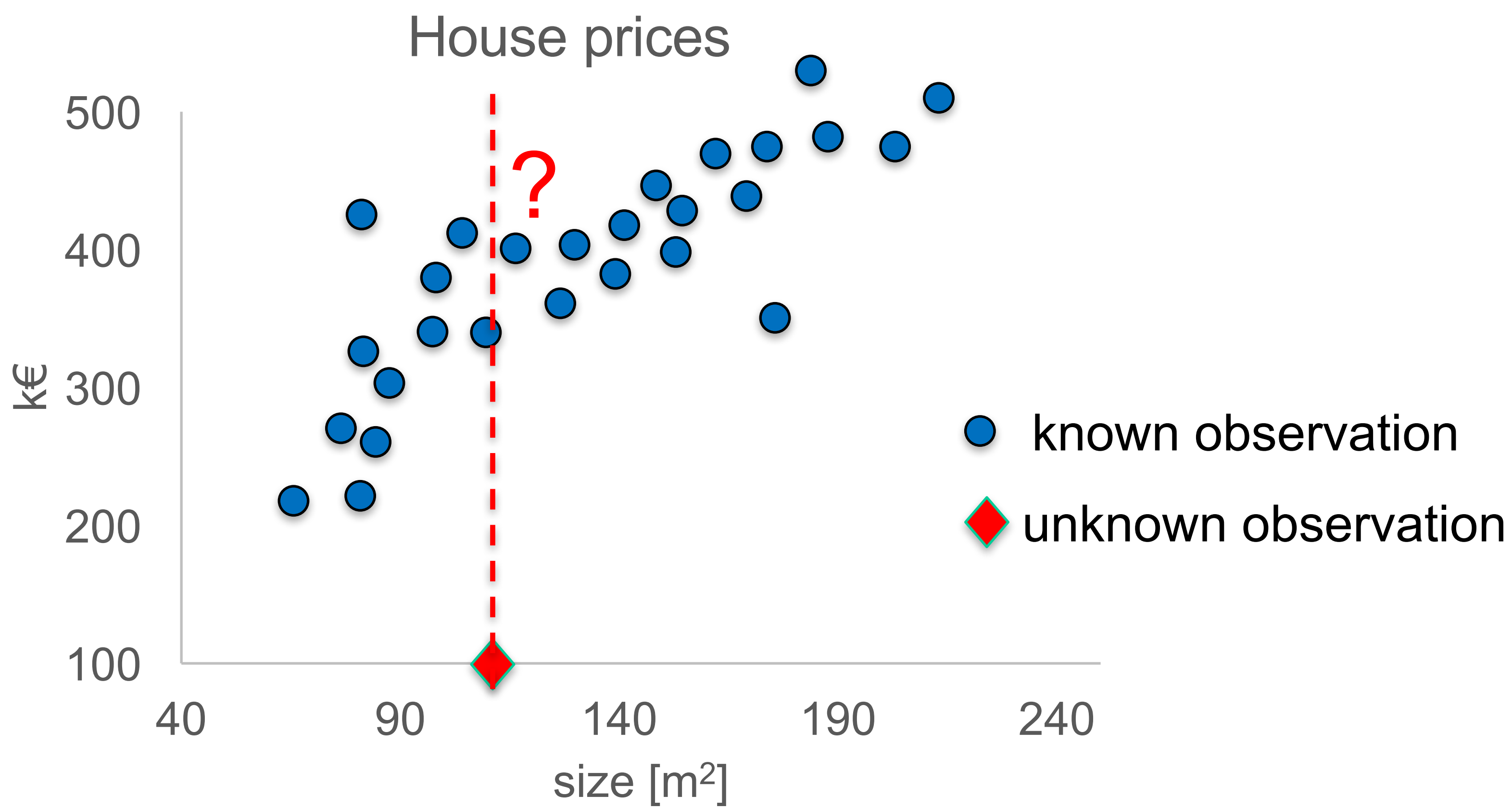
Reinforcement Learning: the algorithm learns by receiving feedback on the effect of modifying some parameters, e.g. the power and the modulation

Some basic concepts in ML

- Data shall be manipulated with care!
  - Overfitting vs Underfitting
  - Linear vs Non Linear models

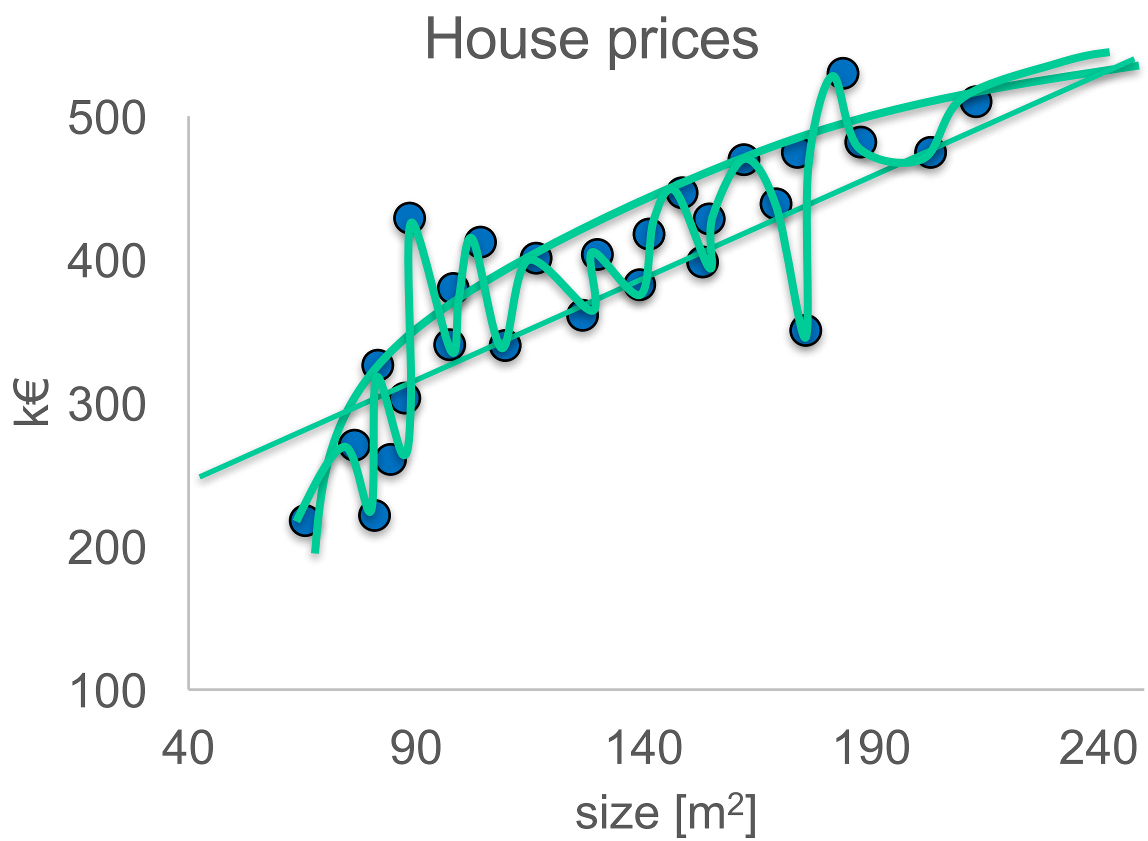
An example of fitting a model

- Suppose we want to predict a house price given its size



An example of fitting a model

- Linear, quadratic, polynomial, non linear



SIDE MESSAGE:

Increasing model flexibility might lead to overfitting!!!

Predictions can get worse if the model is too flexible (counter-intuitive!)

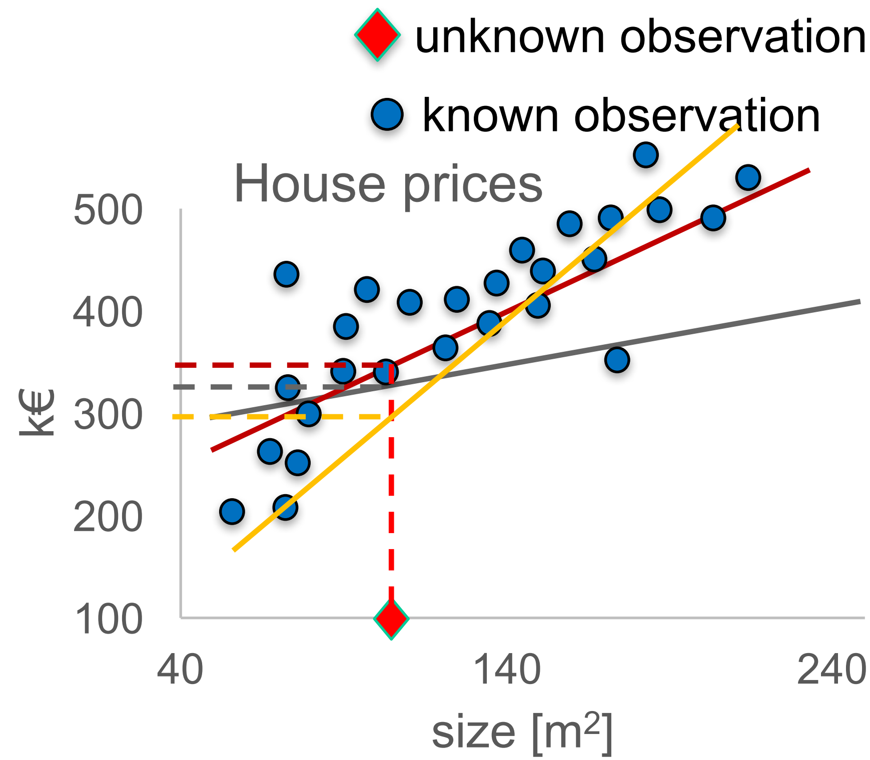
Some algorithms

- Supervised
  - Parametric
    - Linear regression ..
    - Logistic regression
    - Neural Networks
    - SVM
  - Non parametric
    - K-nearest neighbor
    - Random Forest
- Unsupervised
  - Clustering
    - K-means
    - Gaussian Mixture Models

- Simplest model
  - *h*($\underline{x}$) is a ***linear*** function
  - *h*($\underline{x}$) has only **one variable** (univariate), i.e., feature x_1

$h(\underline{x}) = h(x_1) = \theta_0 + \theta_1 x_1$

  - θ_0 and θ_1 are the "weights"
- How to choose θ_0 and θ_1 ?



Minimize the training mean-square error (MSE)

$$\min_{\theta_0,\theta_1}\left\{MSE = \frac{1}{2m}\sum_{i=1}^{m}\left(h(x^{(i)}) - y^{(i)}\right)^2\right\}$$

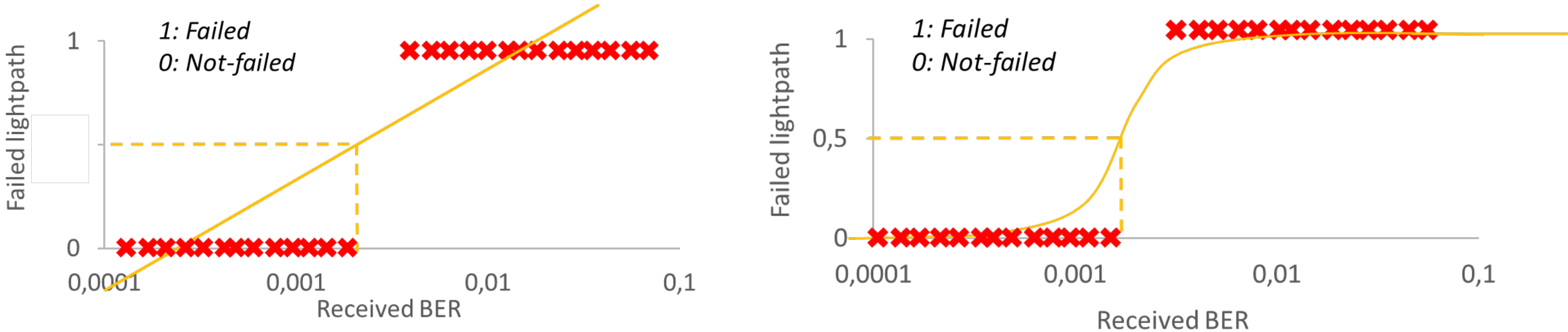
Multivariate and Polynomial Regression

- Multivariate
  - We now have a features **vector** $\underline{x}=(x_1, x_2, \ldots x_n)$
  - $h(\underline{x}) = h(x_1 \ldots x_N) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots\ldots \theta_N x_N$
    - $\theta_0 \ldots \theta_1$ are the "weights" chosen by the algorithm

- Polynomial
  - We now increasing the order of polynomials in $h(x)$
  - $h(\underline{x}) = \theta_0 + \theta_1 x_1 + \theta_{12} x_1 x_2 + \theta_2 (x_2)^2 \ldots + \theta_n x_n$

Logistic regression [Classification!]

- Output *h(x)* takes only **discrete** values
  - Ex: y={0;1}, e.g., yes/no, good/bad, spam/non-spam…
  - Multiclass classifier: y={A,B,C,…}

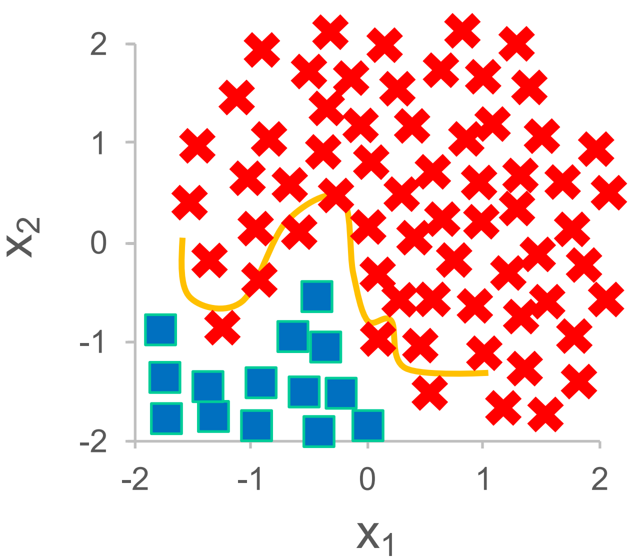


- A good candidate function *h* for
  - *h(z) = 1/(1+e^-z^)* is the "logistic" (or "sigmoid") function
    - for z → -inf: h(z)→0
    - for z → +inf: h(z)→1
    - for z=0: h(z)=0.5

Neural Networks

- Why do we need a new algorithm?
  - Some problems are just too complex
    - Many features can have a role → increased features space
  - Difficult for a human to even know which features are important

$$h(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \ldots)$$



Suppose we have 100 different features and we want to add all quadratic terms:

$x_1^2,\ x_1x_2,\ \ldots,\ x_1x_{100}\ ;$

$x_2^2,\ \ldots,\ x_2x_{100}\ ;$

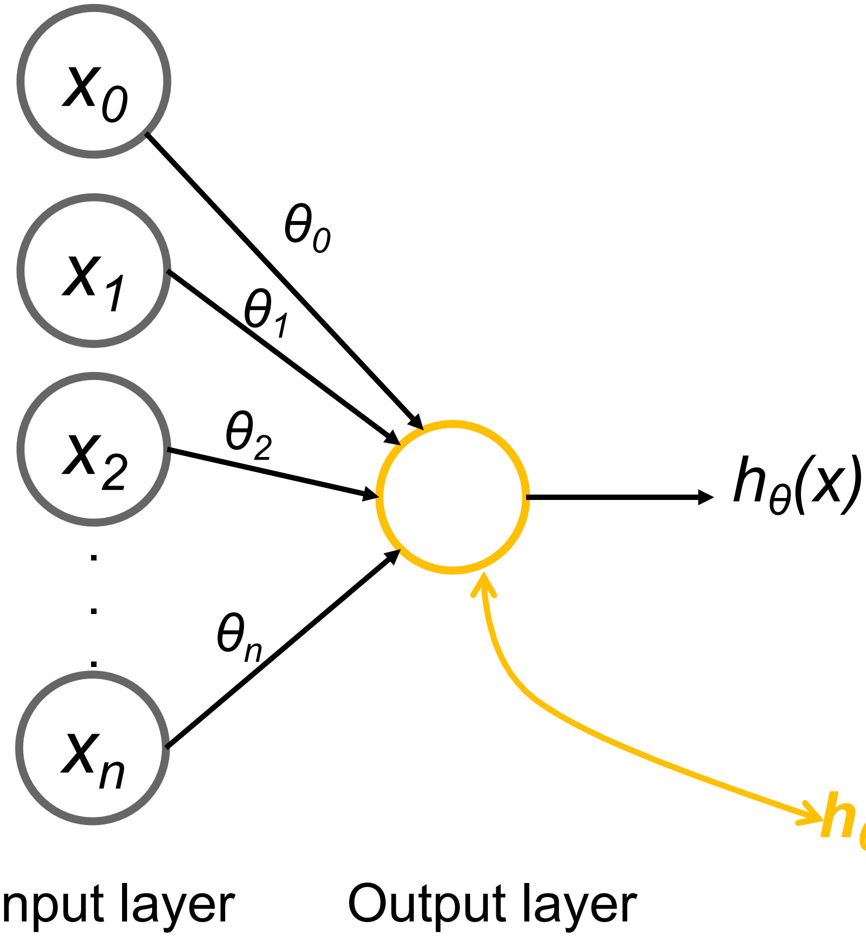
$\ldots$

$x_{99}^2,\ x_{99}x_{100}\ ;$

$x_{100}^2.$

n "original" features require $O(n^2)$ quadratic terms!

- The simplest neural network



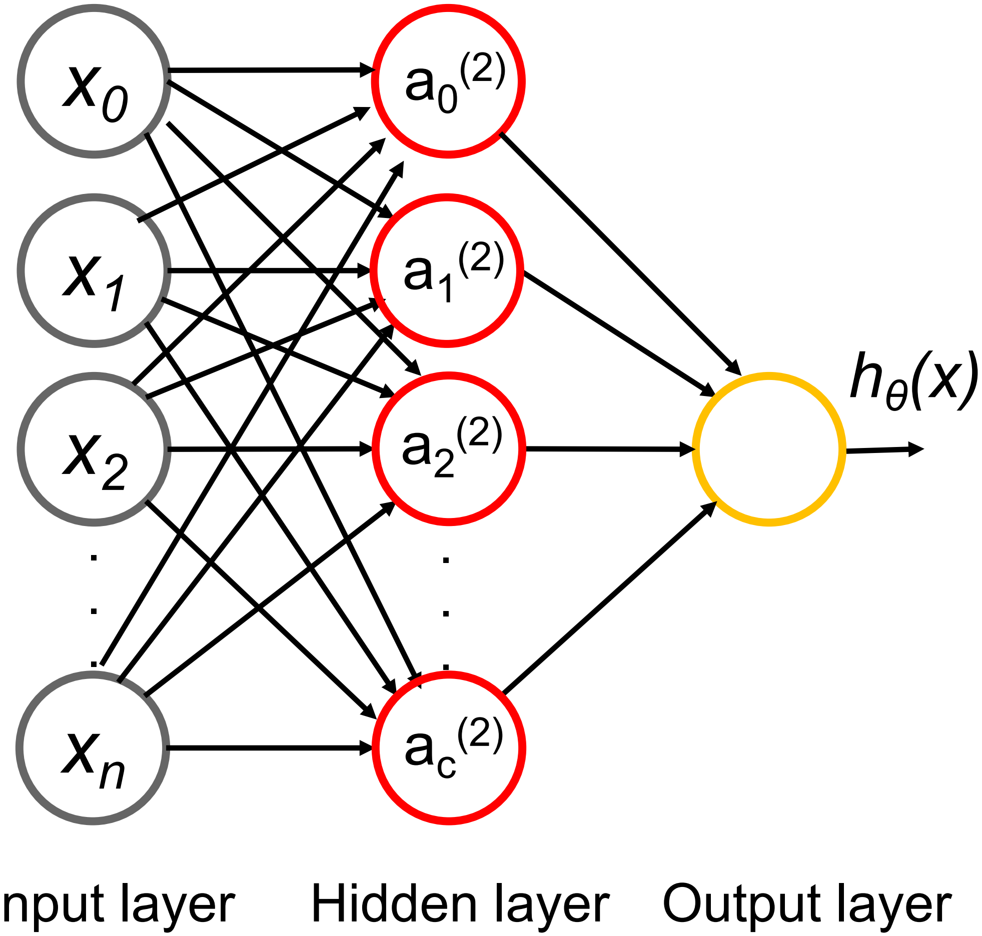
$$x = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \qquad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \dots \\ \theta_n \end{bmatrix}$$

inputs or features — weights

$$h_\theta(x) = g(\theta^T x) = 1/(1+e^{-(\theta^T x)})$$

Neural Networks (NN) representation
Multiple layers

- A “collection” of interacting neurons

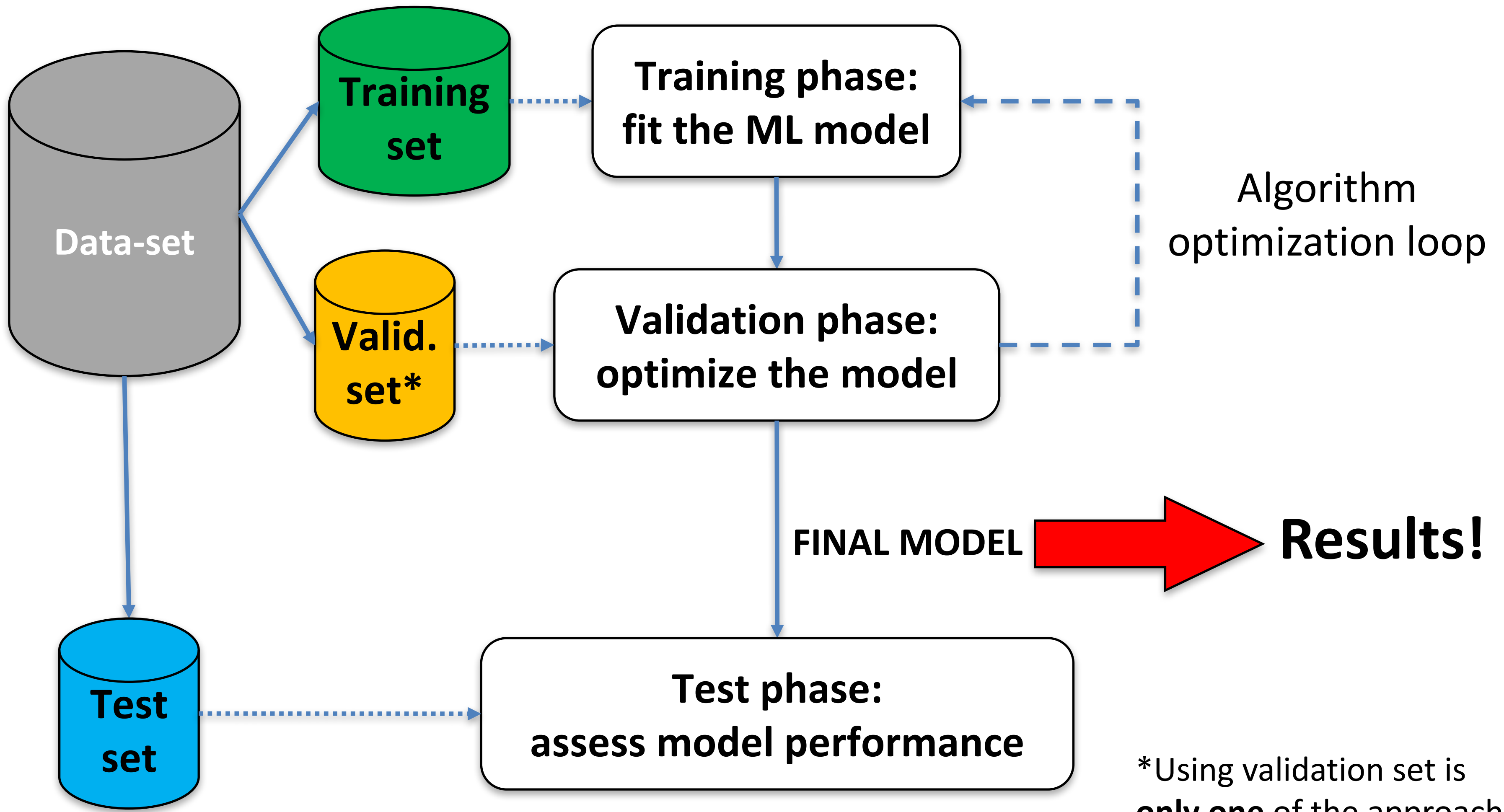


3 observations:

1. NN can capture any relation between x and y
2. Hyperparameters: # of hidden layers , # neurons per hidden layer.
Who decides them?
3. Deep Learning: the more Layer, the less decisions shal be taken by a programmer

Use of data in Machine Learning

Training, testing, validation

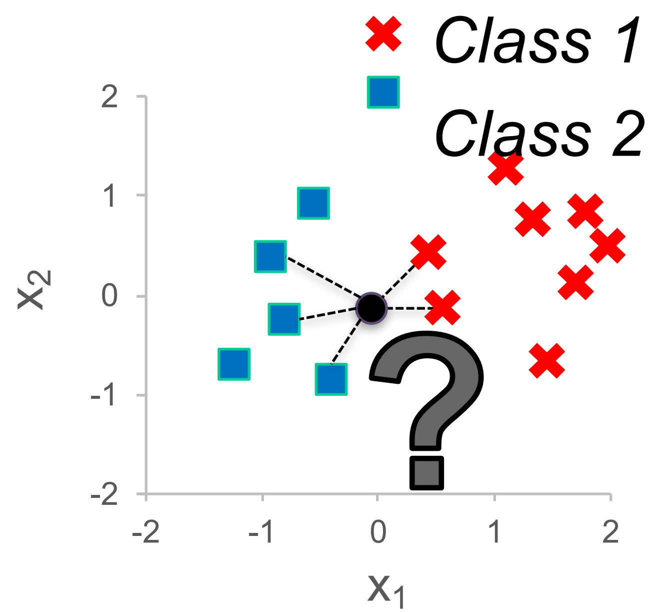


*Using validation set is **only one** of the approaches to optimize a ML algorithm

K-Nearest Neighbors

- Used for classification and regression
- Decision based on the K nearest points in the training sets
  - Need to choose K

- Example 1: classification (K=3)
  - Choose the most frequent class among the KNN → predict class 1
  - Changing the value of K (e.g. K=5) may affect the result → predict class 2



Performance metrics in Machine Learning

- Let us focus only on classifiers
  - Note: the output produced by the classifier is P_{pos}
    - Ideally, we want P_{pos} close to 1

- Metrics used in the following
    - Accuracy
      - Fraction of test instances correctly classified
      - Affected by by relative frequency
    - Area under the ROC curve
- Other metrics
    - Precision
    - Recall

Several questions to be addressed

- Which ML algorithm best describes our problem?
- Which data/features should we consider to make predictions?
- Is it worth collecting as much data as possible? Is there any irrelevant parameter we can (or should) neglect?
- What is the performance of our learning algorithm?
- And what is its complexity?
- …

Outline: Part 2 Applications

1. ML for QoT Estimation for Unestablished Lighpaths
   - L. Barletta, A. Giusti, C. Rottondi and M. Tornatore, *QoT Estimation for Unestablished Lighpaths using Machine Learning*, OFC 2017, paper Th1J.1
   - C. Rottondi, L. Barletta, A. Giusti and M. Tornatore, *A Machine Learning Method for Quality of Transmission Estimation of Unestablished Lightpaths*, in IEEE/OSA Journal of Optical Comm.& Netw. Vol. 10, No. 2, Feb. 2018
2. ML for Soft-Failure Identification
   - S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks," in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2018
   - A. Vela et al., "BER degradation Detection and Failure Identification in Elastic Optical Networks", in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017
3. An overview of other applications at network layer
   - F. Musumeci et al., "A Survey on Application of Machine Learning Techniques in Optical Networks", Submitted to IEEE communication surveys and tutorials, available in ArXiv
   - Javier Mata, et a., Artificial intelligence (AI) methods in optical networks: A comprehensive survey, Optical Switching and Networking, Volume 28, 2018, pp. 43-57

Outline: Part 2
Applications

1. ML for QoT Estimation for Unestablished Lighpaths

- L. Barletta, A. Giusti, C. Rottondi and M. Tornatore, *QoT Estimation for Unestablished Lighpaths using Machine Learning*, OFC 2017, paper Th1J.1
- C. Rottondi, L. Barletta, A. Giusti and M. Tornatore, *A Machine Learning Method for Quality of Transmission Estimation of Unestablished Lightpaths*, in IEEE/OSA Journal of Optical Comm.& Netw. Vol. 10, No. 2, Feb. 2018

2. ML for Soft-Failure Identification

- S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, “Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks,“ in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2018
- A. Vela et al., “BER degradation Detection and Failure Identification in Elastic Optical Networks”, in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017

3. An overview of other applications at network layer

- F. Musumeci et al., “A Survey on Application of Machine Learning Techniques in Optical Networks”, Submitted to IEEE communication surveys and tutorials, available in ArXiv
- Javier Mata, et a., Artificial intelligence (AI) methods in optical networks: A comprehensive survey, Optical Switching and Networking, Volume 28, 2018, pp. 43-57

Motivation
Increasing «degrees of freedom»

- A wider range of **degrees of freedom** (parameters) is available to system engineers:
  - path
  - spectrum
  - modulation format
  - baud rate
  - FEC coding
  - single/multicarrier transmission
  - nonlinearity mitigation solution
  - adaptive channel spacing
  - …
- Combinations of these lighpath parameters grow dramatically
- Possibly, for all of these combinations, we shall calculate a QoT

Existing (pre-deployment) estimation techniques for lightpath QoT

- **"Exact" analytical models** estimating physical layer impairments (e.g., split-step Fourier method…)
  - Accurate results
  - Heavy computational requirements
  - Not scalable to large networks and real time estimations

- **Marginated formulas** (Power Budget, Gaussian model…)
  - Faster and more scalable
  - Inaccurate, high margination, underutilization of network resources (up to extra 2 dB for design margins [1])

[1] Y. Pointurier, "Design of low-margin optical networks," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9-A17, Jan. 2017. doi: 10.1364/JOCN.9.0000A9

Machine Learning as an alternative approach?

- Machine Learning exploits knowledge extracted from field data…
  - QoT of established lightpaths, e.g. using monitors (OPMs) at the receiver
- …. to predict the QoT of unestablished lightpaths

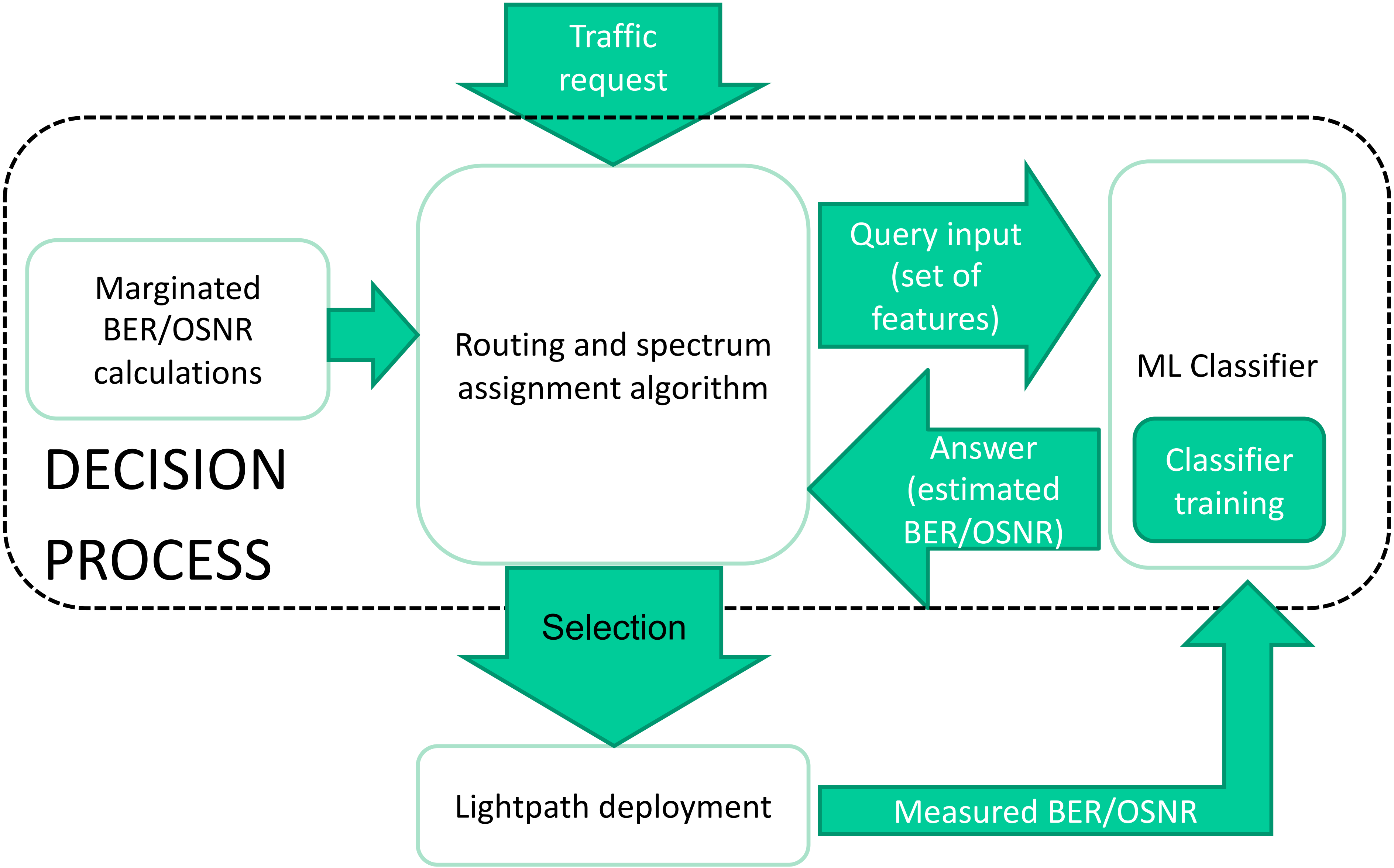
🙂 No need for complex analytical models

🙂 Fast and scalable

🙁 Requires training phase with historical data

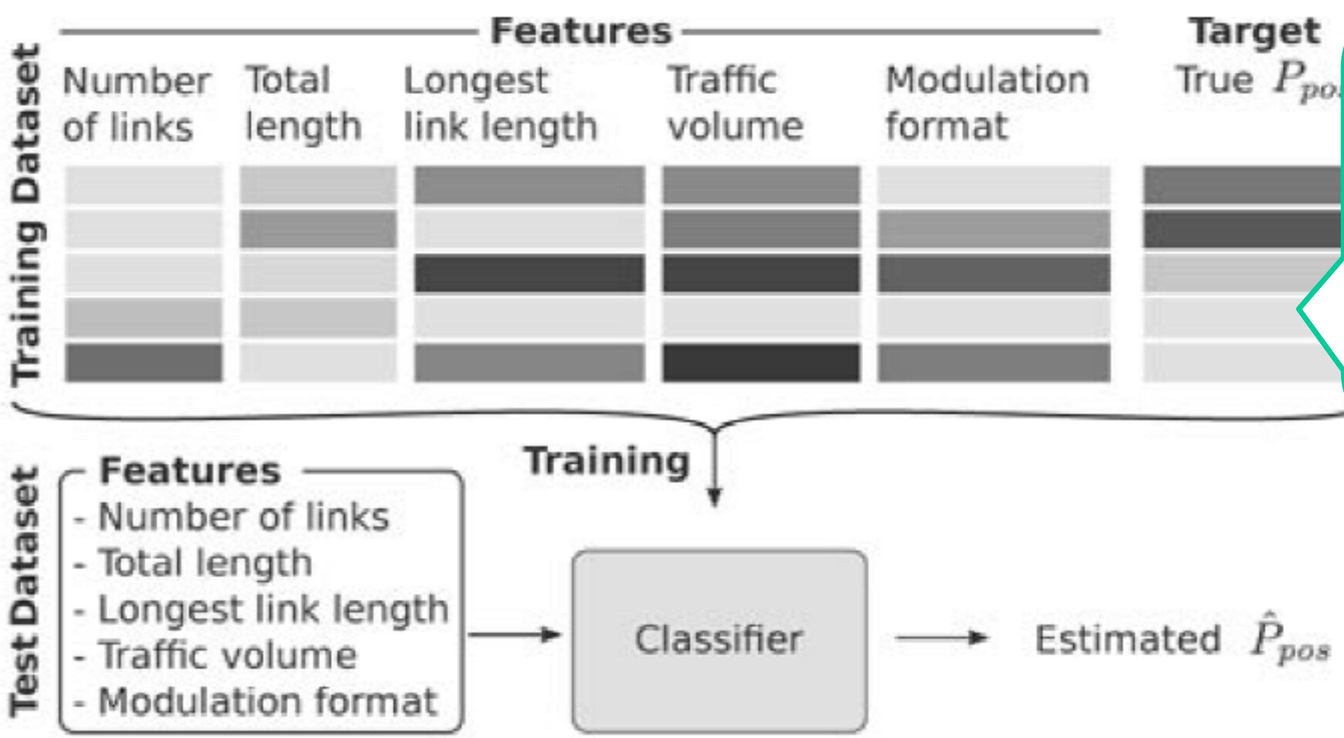
- How long must the training phase be?
- How accurate will the estimation be?
- Objectives of our numerical analysis….

Whole framework: RSA + QoT estimation



How our proposed ML classifier works
Case 1

Output: probabilty that BER ≤ T*
Input: set of lightpath features



The classifier is trained on a set of *L* experiments to generate the ***ground truth***

(Case of only local knowledge)

How our proposed ML classifier works
Case 2

- To the previous 6 feature we add, for the «most interfereing left and right neighbors»:
  - guardband
  - traffic volume
  - modulation format

(Case of complete knowledge)

- Note: these additional six features are chosen with the intent to capture cross-channel nonlinear effects

How to generate synthetic field data?

We use a Bit Error Rate Estimation Tool (ETool) that on input of …

- a candidate lightpath
- a modulation format

.. and under assumption of..

- AWGN channel
- back-to-back penalties
- a random system margin expneg distributed with mean 2 dB
  - Expneg provides a worst case

… calculates …

- BER measured at the input of the channel decoder

Which Machine Learning Algorithm?

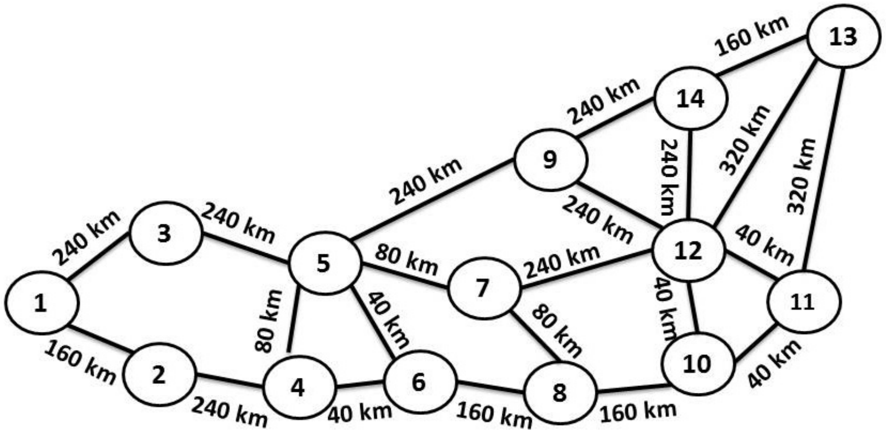
- We use a Random Forest (RF) classifier with 25 estimators
- To take this choice, we have compared:
  - 5 RF classifiers
  - 3 k-Nearest-Neighbor classifiers

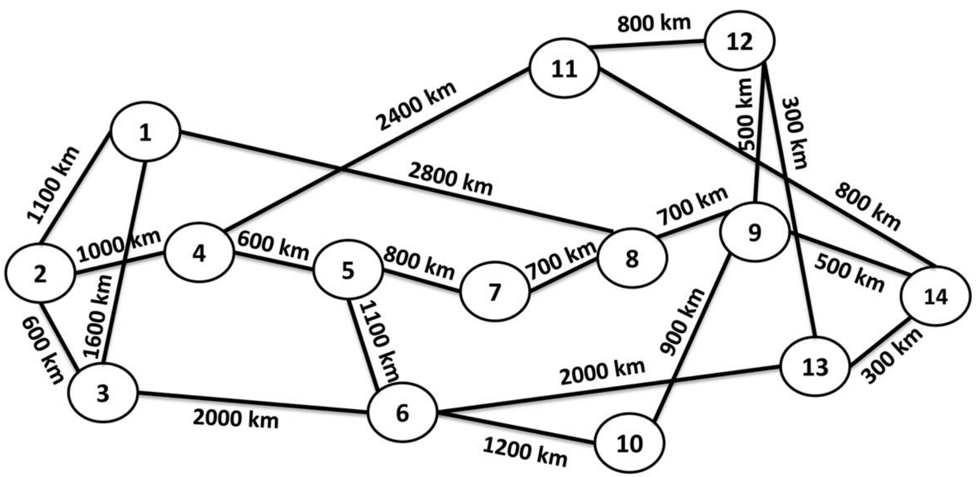
Algorithm	Training time (s)	Test time (s)	AUC	Accuracy
Dummy classifier	0.048979	3.83 e-07	0.501	0.539
1 Nearest Neighbor	1.183121	4.83 e-05	0.959	0.957
5 Nearest Neighbor	1.085116	5.05 e-05	0.991	0.965
25 Nearest Neighbor	1.211694	6.91 e-05	0.996	0.965
Random Forest 1 tree	0.076944	3.96 e-07	0.991	0.965
Random Forest 5 trees	0.180835	6.24 e-07	0.995	0.970
Random Forest 25 trees	0.721042	1.56 e-06	0.996	0.968
Random Forest 100 trees	2.830545	5.32 e-06	0.996	0.966
Random Forest 500 trees	14.052182	2.63 e-05	0.996	0.966

- RF with 25 estimators provided the best trade-off between performance and computational time

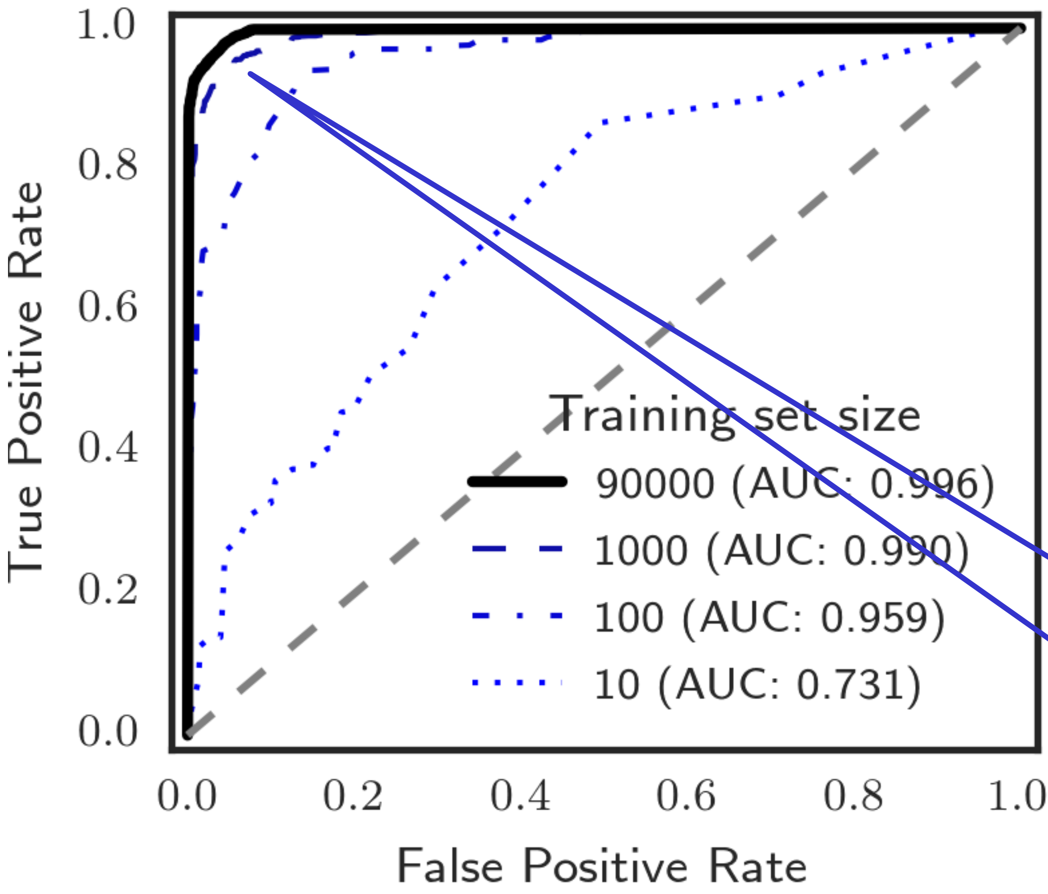
Training and Testing Scenario

- Japanese and NSF optical network
- Flexgrid @ 12.5 GHz
- Transceivers @ 28 GBaud
- 6 Modulation formats
  - (DP) BPSK, QPSK, 8-QAM to 64-QAM,
- Traffic requests: [50;1000] Gbps
- 3 candidate paths per node pair
- BER threshold $T = 4*10^{-3}$





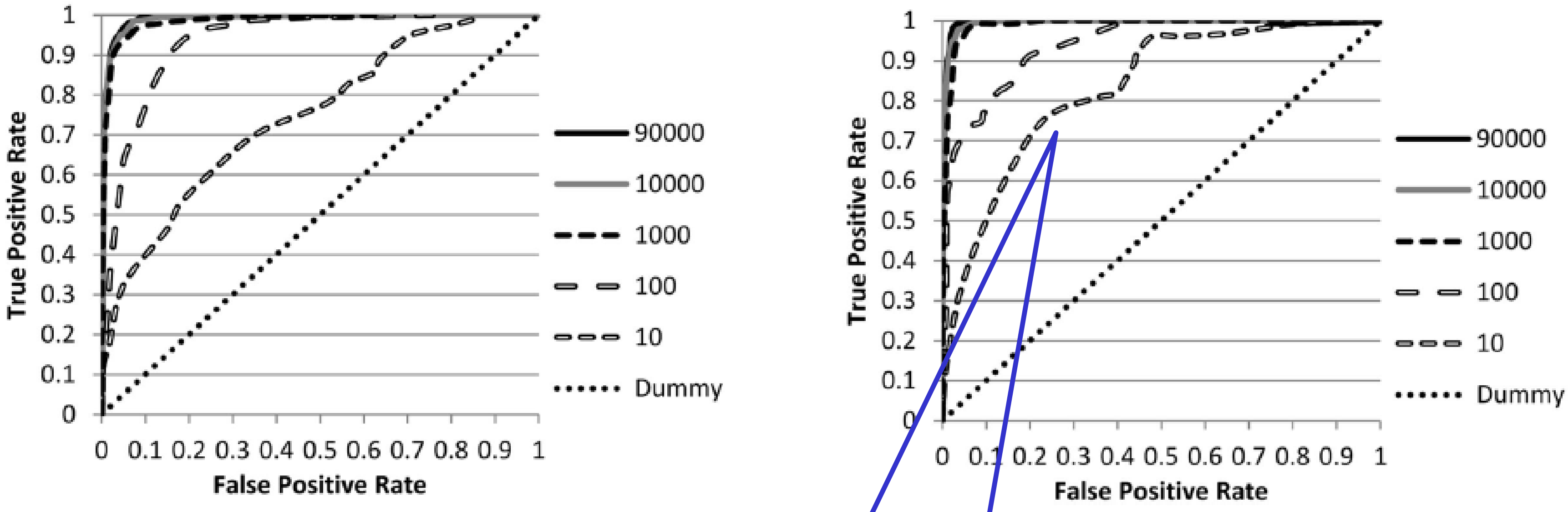
How long shall training phase be?
(1) *Accuracy vs training set size*



- «ROC» curve
- Area under the ROC curve (AUC)

Take-Away 1: Training phase has a reasonable duration

How long shall training phase be? (2) *Effect of topology on Accuracy*



(a) Japan topology

(b) NSF topology

Take-Away 2: easier to classify on a large network (less options!)

How to build the training dataset?

- Use historical data
  - We will never observe samples of with too high BER!!
- Use random probes:
  - Very costly (high spectrum occupation)
- Use selective probes:
  - Lower spectrum occupation, good accuracy

TABLE V: AUC comparison of probing approaches

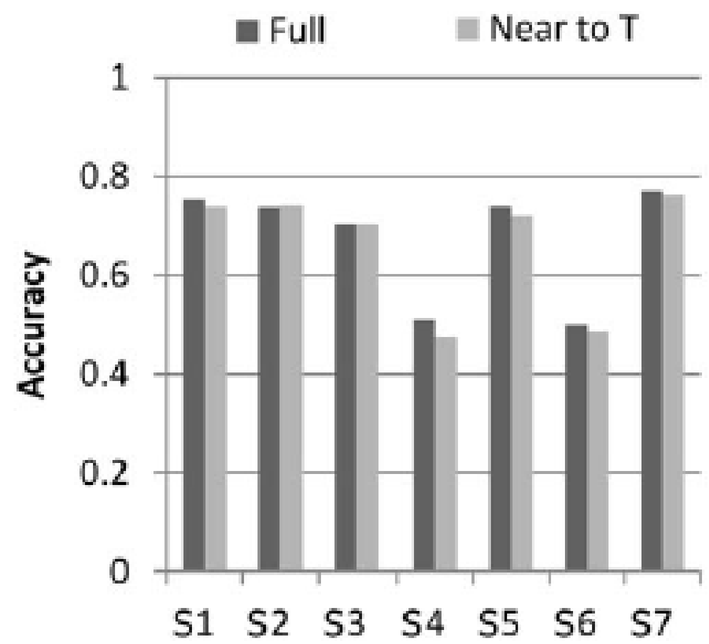
Training set	AUC (full testing dataset)
C (historical)	0.77
C (selective, 5% probes)	0.85
C (selective, 10% probes)	0.87
C (selective, 25% probes)	0.89
C (selective, 50% probes)	0.89
A (random)	0.98

Analysis of feature relevance

- Removing irrelevant «ML-input features» makes the system less costly and less complex to manage

TABLE IV: The considered feature subsets

	S1	S2	S3	S4	S5	S6	S7
number of links	✓	✓	✓	✓			
lighpath length	✓	✓	✓	✓	✓	✓	
length of longest link	✓	✓	✓	✓			
traffic volume	✓	✓	✓		✓		✓
modulation format	✓	✓		✓	✓	✓	✓
guardband, modulation format and traffic volume of nearest left and right neighbor	✓						



A note on resource consumption

- Ok, by automating QoT estimation we can save workforce/Opex and decrease setup times. But, what is the impact on resource saving?

Output: probabilty that BER ≤ T*

γ: threshold you are willing to accept

γ	Average reduction in number of installed transceivers [%]	Average reduction in overall spectrum occupation [%]
0.5	17.14	16.70
0.6	15.84	15.44
0.7	14.51	13.88
0.8	11.41	10.73
0.9	6.90	6.31
1	1.17	0.96

Outline: Part 2 Applications

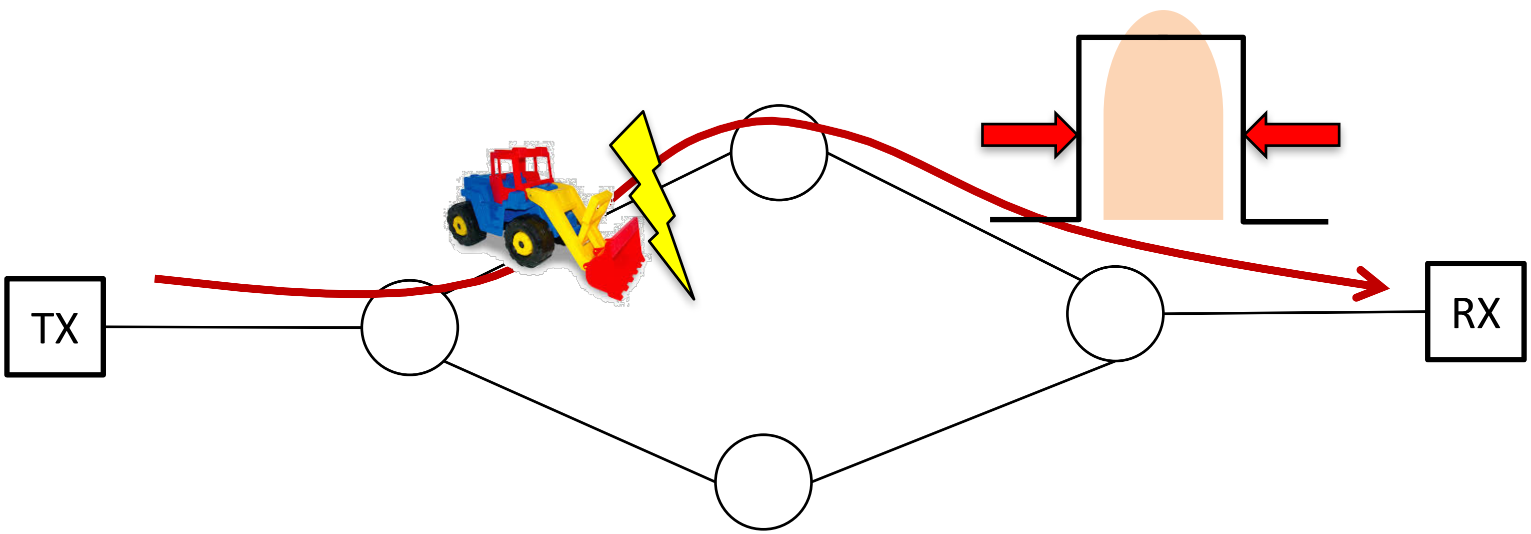
1. ML for QoT Estimation for Unestablished Lighpaths
   - L. Barletta, A. Giusti, C. Rottondi and M. Tornatore, *QoT Estimation for Unestablished Lighpaths using Machine Learning*, OFC 2017, paper Th1J.1
   - C. Rottondi, L. Barletta, A. Giusti and M. Tornatore, *A Machine Learning Method for Quality of Transmission Estimation of Unestablished Lightpaths*, to appear JOCN2018

2. ML for Soft-Failure Identification
   - S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks,“ in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2018
   - Vela et al., "BER degradation Detection and Failure Identification in Elastic Optical Networks", in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017

3. An overview of other applications at network layer
   - F. Musumeci et al., "A Survey on Application of Machine Learning Techniques in Optical Networks", Submitted to IEEE communication surveys and tutorials, available in ArXiv
   - Javier Mata, et a., Artificial intelligence (AI) methods in optical networks: A comprehensive survey, Optical Switching and Networking, Volume 28, 2018, pp. 43-57

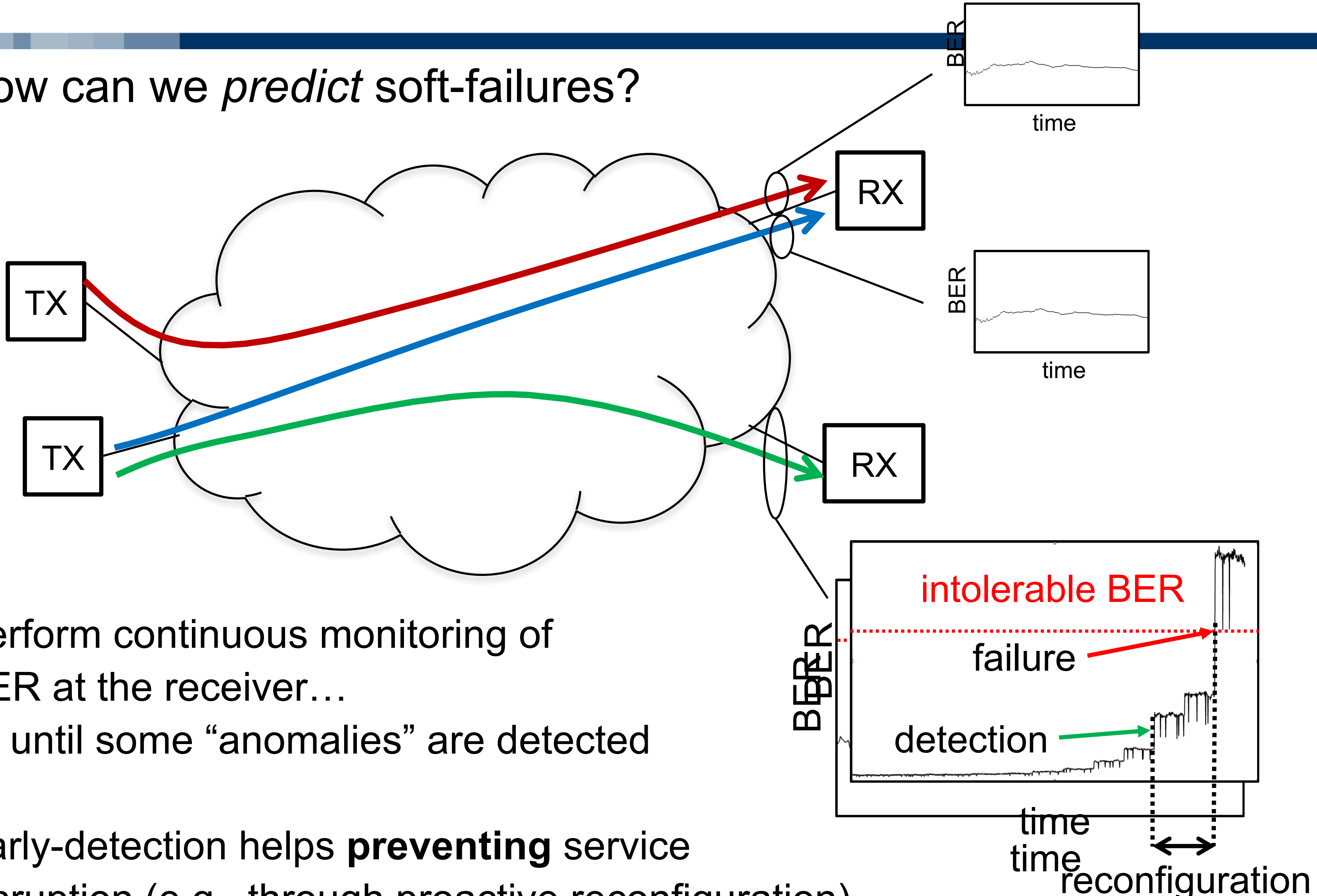
Two main failure types in optical networks

- Hard-failures
  - Sudden events, e.g., fiber cuts, power outages, etc.
  - Unpredictable, require «protection» (*reactive procedures*)
- Soft-failures:
  - Gradual transmission degradation due to equipment malfunctioning, filter shrinking/misalignment…
  - Trigger early network reconfiguration (*proactive* procedures)



Soft-failure *early-detection*

- How can we *predict* soft-failures?



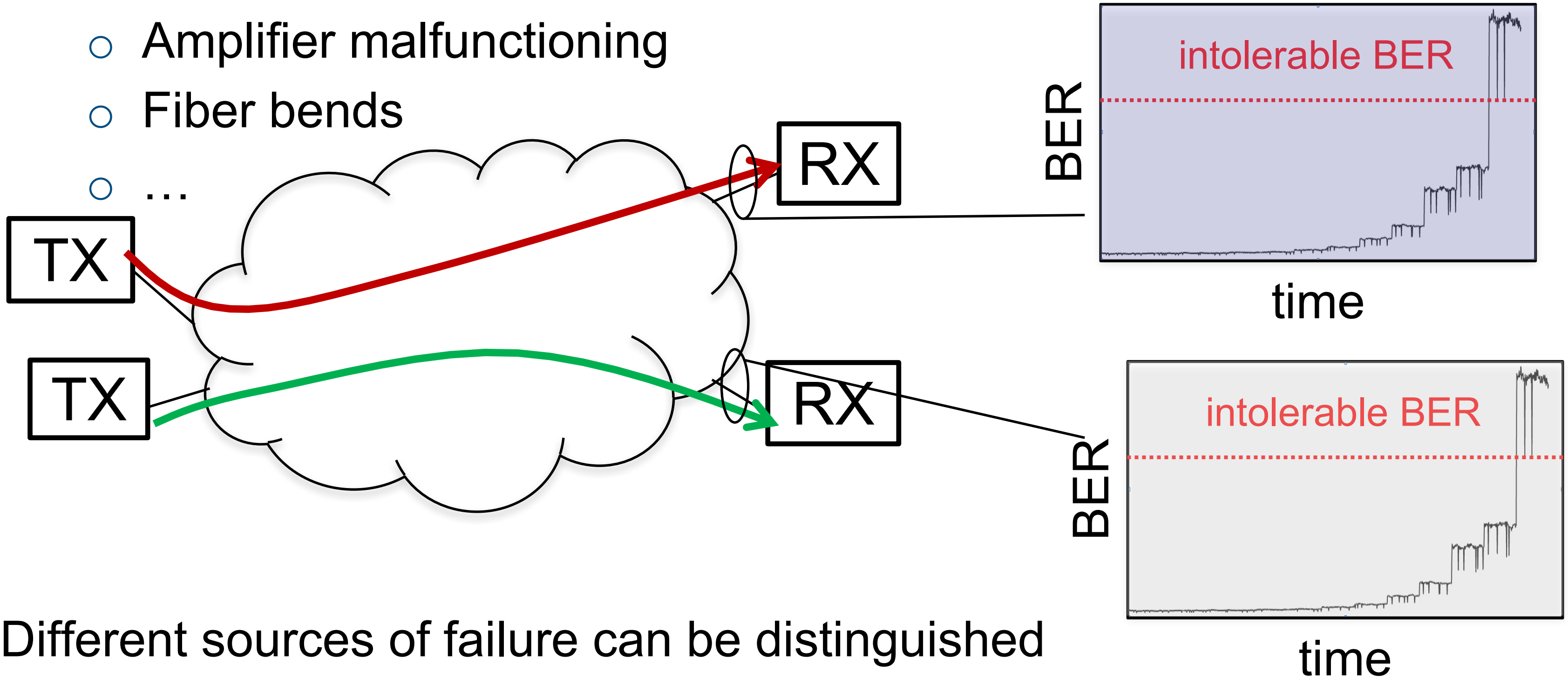
Perform continuous monitoring of BER at the receiver…
… until some "anomalies" are detected

Early-detection helps **preventing** service disruption (e.g., through proactive reconfiguration)

Vela et al., "BER degradation Detection and Failure Identification in Elastic Optical Networks", in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017

Soft-failure *cause identification*

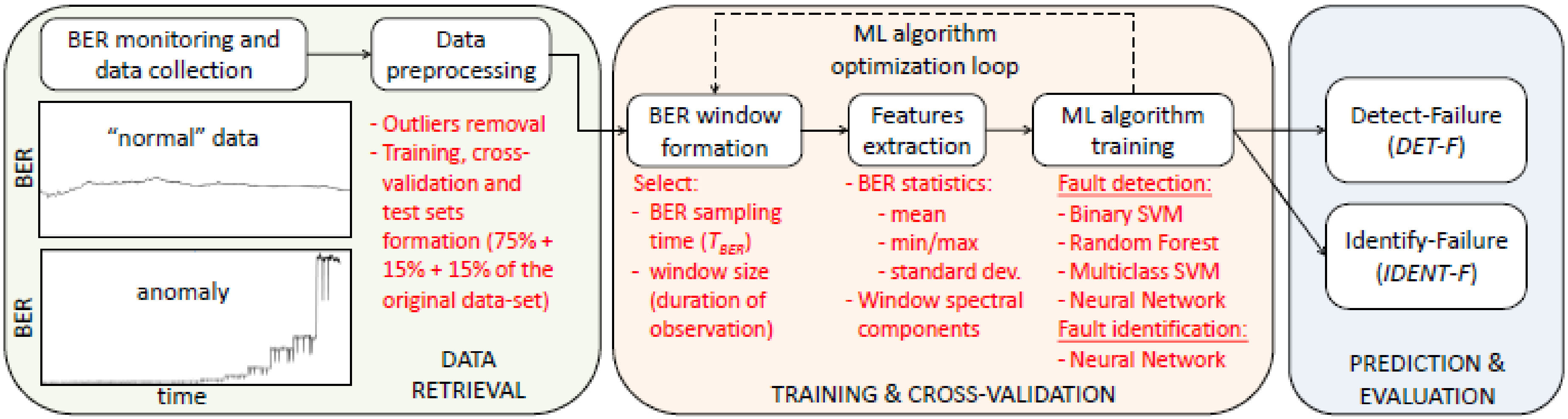
- How can we identify the *cause* of the failure?
  - Failures can be caused by different sources
    - Filters shrinking/misalignment
    - Amplifier malfunctioning
    - Fiber bends
    - …

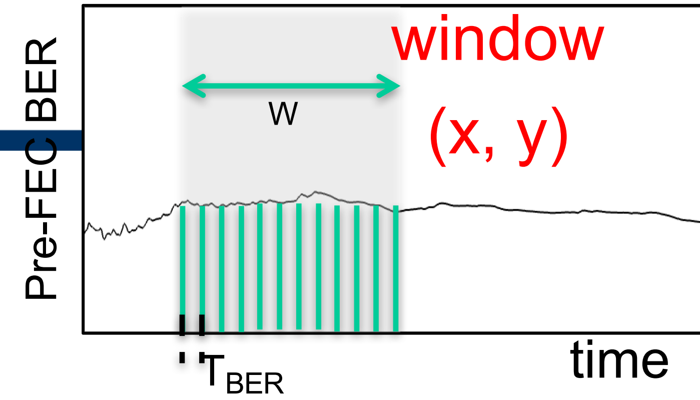


Different sources of failure can be distinguished via the different effects on BER (i.e., via different BER "features")

S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks,"in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2018

Phases of our study
Overall view





1. Data Retrieval

3 decisions

Validation (optimization of hyperparameters)

BER window → Features → ML algorithm → **3. Prediction and Evaluation**

BER window

Select:
- BER sampling time (T_{BER})
- window size (duration of observation)

Features

- BER statistics:
  - mean
  - min/max
  - standard dev.
- Window spectral components
- Feature Scaling

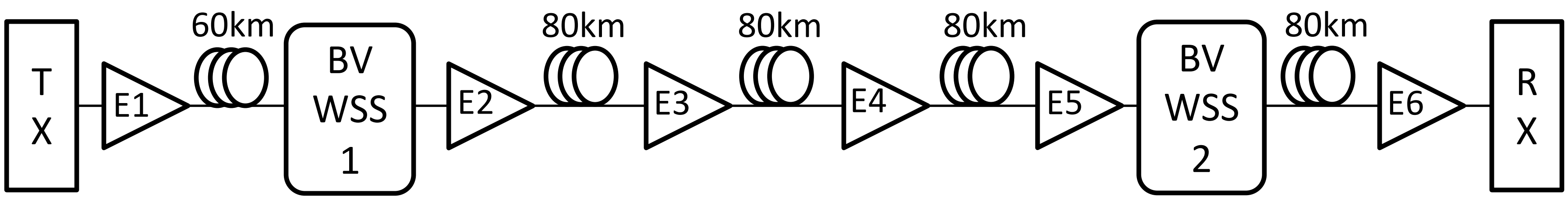
ML algorithm

Fault detection:
- Binary SVM
- Random Forest
- Multiclass SVM
- Neural Network

Fault identification:
- Neural Network

Testbed setup

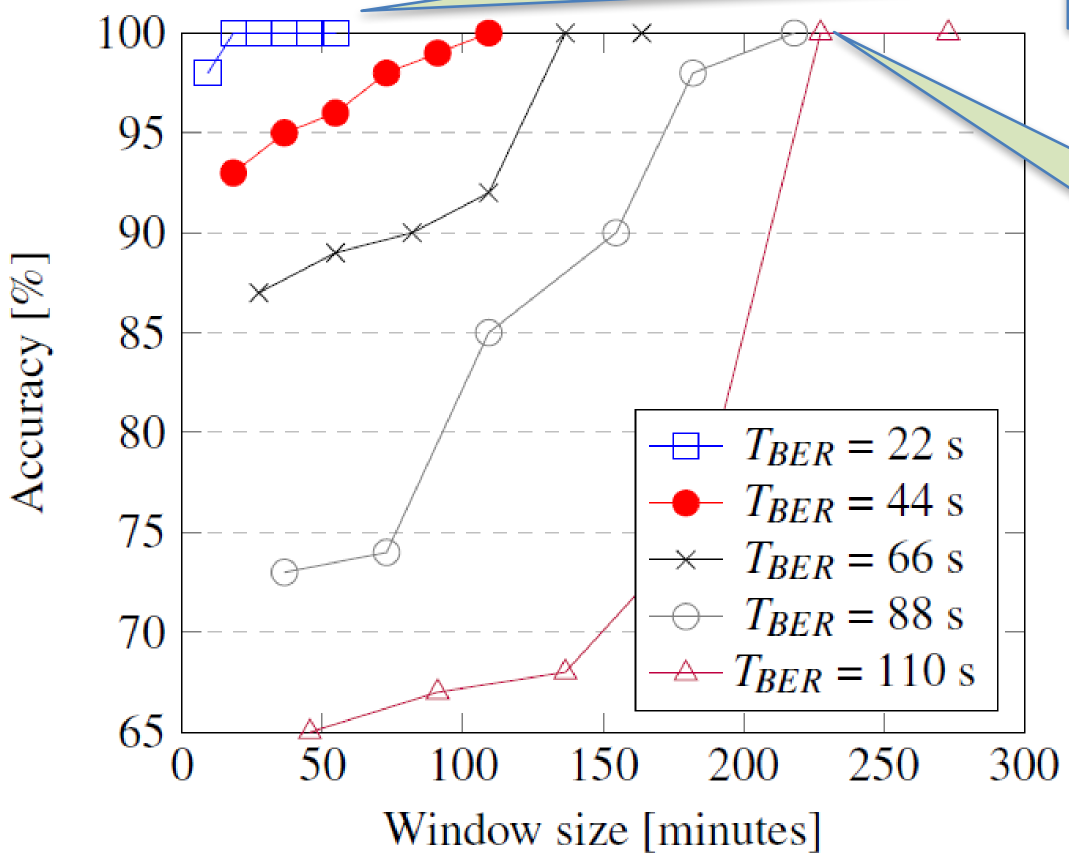
- Testbed for real BER traces
  - Ericsson 380 km transmission system
    - 24 hours BER monitoring
    - 3 seconds sampling interval
  - PM-QPSK modulation @ 100Gb/s
  - 6 Erbium Doped Fiber Amplifiers (EDFA) followed by Variable Optical Attenuators (VOAs)
  - Bandwidth-Variable Wavelength Selective Switch (BV-WSS) is used to emulate 2 types of BER degradation:
    - **Filter misalignment**
    - Additional attenuation in intermediate span (e.g., due to **EDFA gain-reduction**)



Numerical results: *Detection*
Accuracy vs window features

- Binary SVM

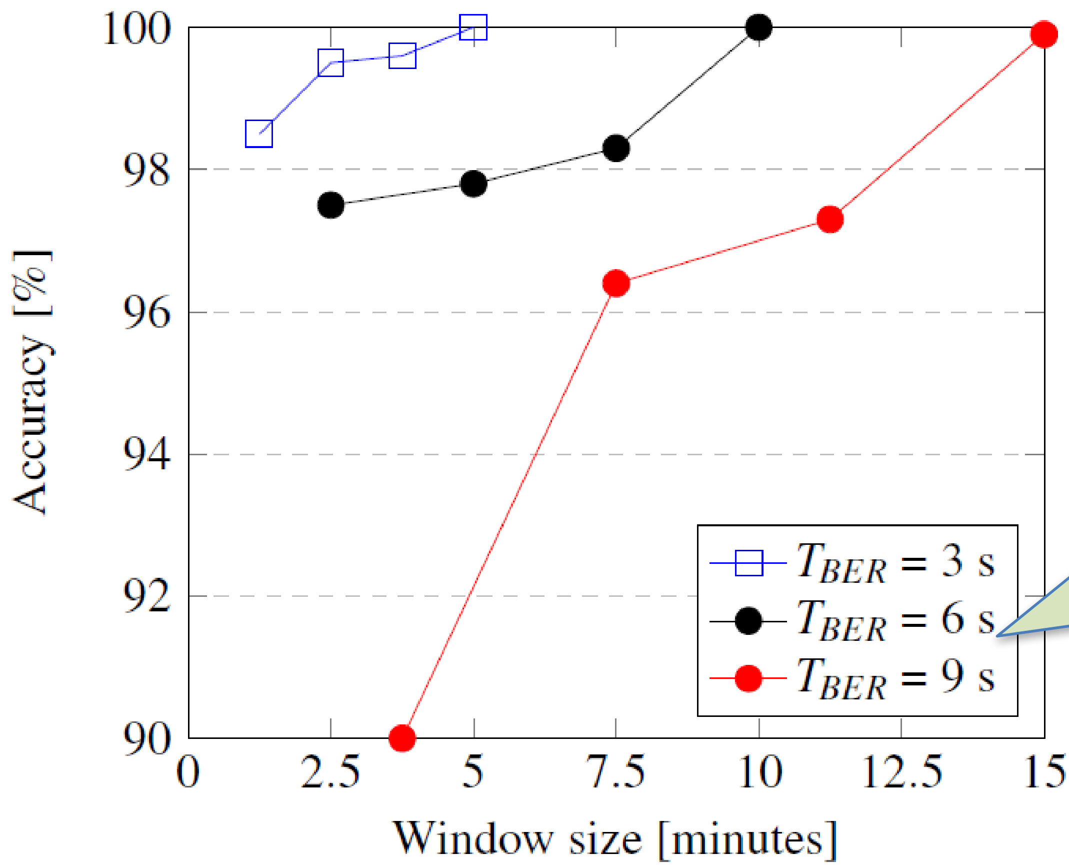
Take-away 1: Higher performance for with low sampling time
→ Fast monitoring equipment is required



Take-away 2: For increasing sampling time, longer "Windows" are needed for high accuracy

Numerical results: *Identification*
Accuracy vs window features

- Neural Network



Take-away 3: To perform failure-cause identification, much smaller sampling period is needed wrt failure detection

Outline: Part 2 Applications

1. ML for QoT Estimation for Unestablished Lighpaths
   - L. Barletta, A. Giusti, C. Rottondi and M. Tornatore, *QoT Estimation for Unestablished Lighpaths using Machine Learning*, OFC 2017, paper Th1J.1
   - C. Rottondi, L. Barletta, A. Giusti and M. Tornatore, *A Machine Learning Method for Quality of Transmission Estimation of Unestablished Lightpaths*, to appear JOCN2018
2. ML for Soft-Failure Identification
   - S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, "Machine-Learning-Based Soft-Failure Detection and Identification in Optical Networks," in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2018
   - Vela et al., "BER degradation Detection and Failure Identification in Elastic Optical Networks", in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017
3. An overview of other applications at network layer
   - F. Musumeci et al., "A Survey on Application of Machine Learning Techniques in Optical Networks", Submitted to IEEE communication surveys and tutorials, available in ArXiv
   - Javier Mata, et a., Artificial intelligence (AI) methods in optical networks: A comprehensive survey, Optical Switching and Networking, Volume 28, 2018, pp. 43-57

Overview of other applications

Physical layer

1. Quality of Transmission (QoT) estimation
2. Optical amplifier control
3. Modulation format recognition
4. Nonlinearities mitigation

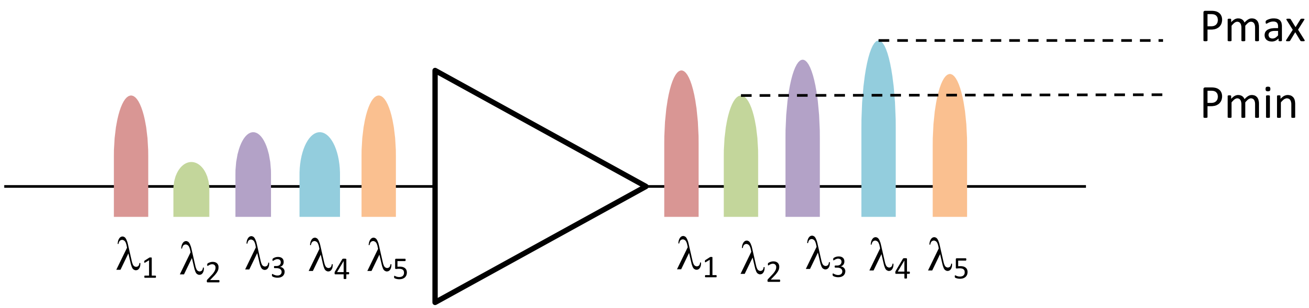
Network layer

1. Traffic prediction and virtual topology design
2. Failure detection and localization
3. Flow classification

Classification taken from: F. Musumeci et al., “A Survey on Application of Machine Learning Techniques in Optical Networks”, Submitted to IEEE Communication Surveys and Tutorials, soon available in ArXiv

Physical layer
Optical amplifier control

- When adding/dropping channels into/from a WDM system, EDFA gain should be adjusted to re-balance output powers

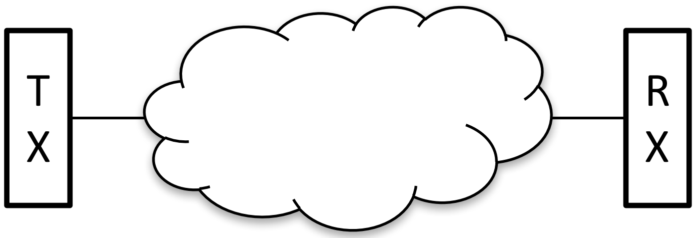


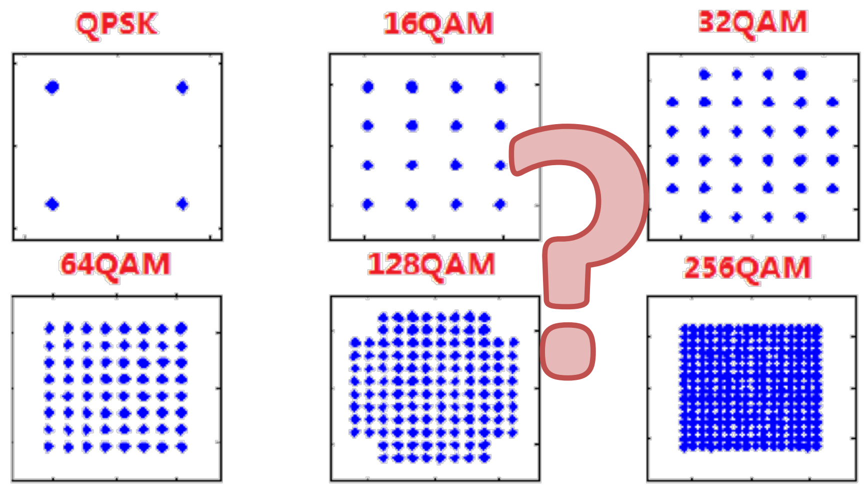
- Analytical models:
  - typically not generalizable
  - depend on the specific system (gain-control mechanism, EDFA gain tilt, nr of EDFAs…) which use to vary during their activity
- ML allows to self-learn typical response patters

Huang et al., "Dynamic mitigation of EDFA power excursions with machine learning", Optics Express, vol. 25 n. 3, Feb. 2017
Bastos et al., "Mapping EDFA Noise Figure and Gain Flatness Over the Power Mask Using Neural Networks", Journal of Microwaves, Optoelectronics and Electromagnetic Applications, vol. 12, n. SI-2, July 2013

Physical layer
Modulation format recognition (MFR)

- Elastic transceiver can to operate with different modulation formats





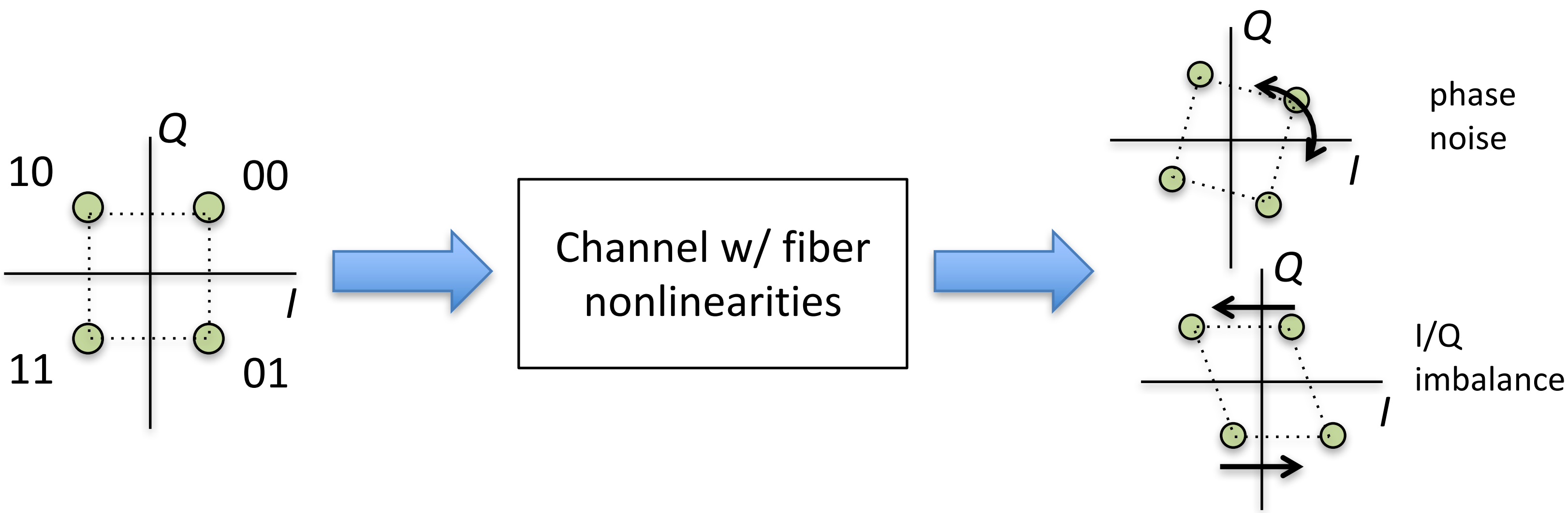
- Traditional MFI requires prior information exchange between end points (from upper layer protocols)
  - additional delay for in signal detection
- ML enables automated MFR from features of the received signal

Khan et al., "Modulation Format Identification in Coherent Receivers Using Deep Machine Learning", Photonics Technology Letters, vol. 28 n. 17, Sep. 2016

Khan et al., "Non-data-aided joint bit-rate and modulation format identification for next-generation heterogeneous optical networks", Optical Fiber Technology, vol. 20 n. 2, Mar. 2014

Tan et al., "Simultaneous Optical Performance Monitoring and Modulation Format/Bit-Rate Identification Using Principal Component Analysis", Journal of Optical Communications and Networking, vol. 6 n. 5, May 2014

- Optical signals are affected by fiber nonlinearities
  - Kerr effect, self-phase modulation (SPM), cross-phase modulation (XPM)…



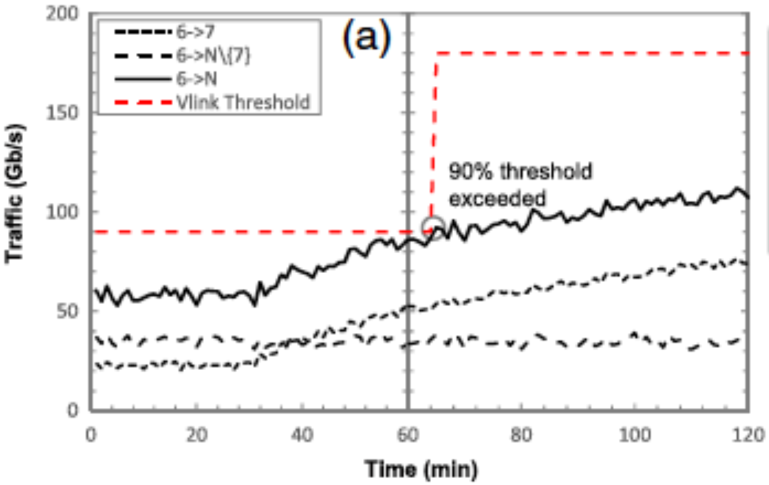
- Traditional methods require complex mathematical models and prior information on the traversed channel
- ML enables "safer" decision by learning from actual channel properties

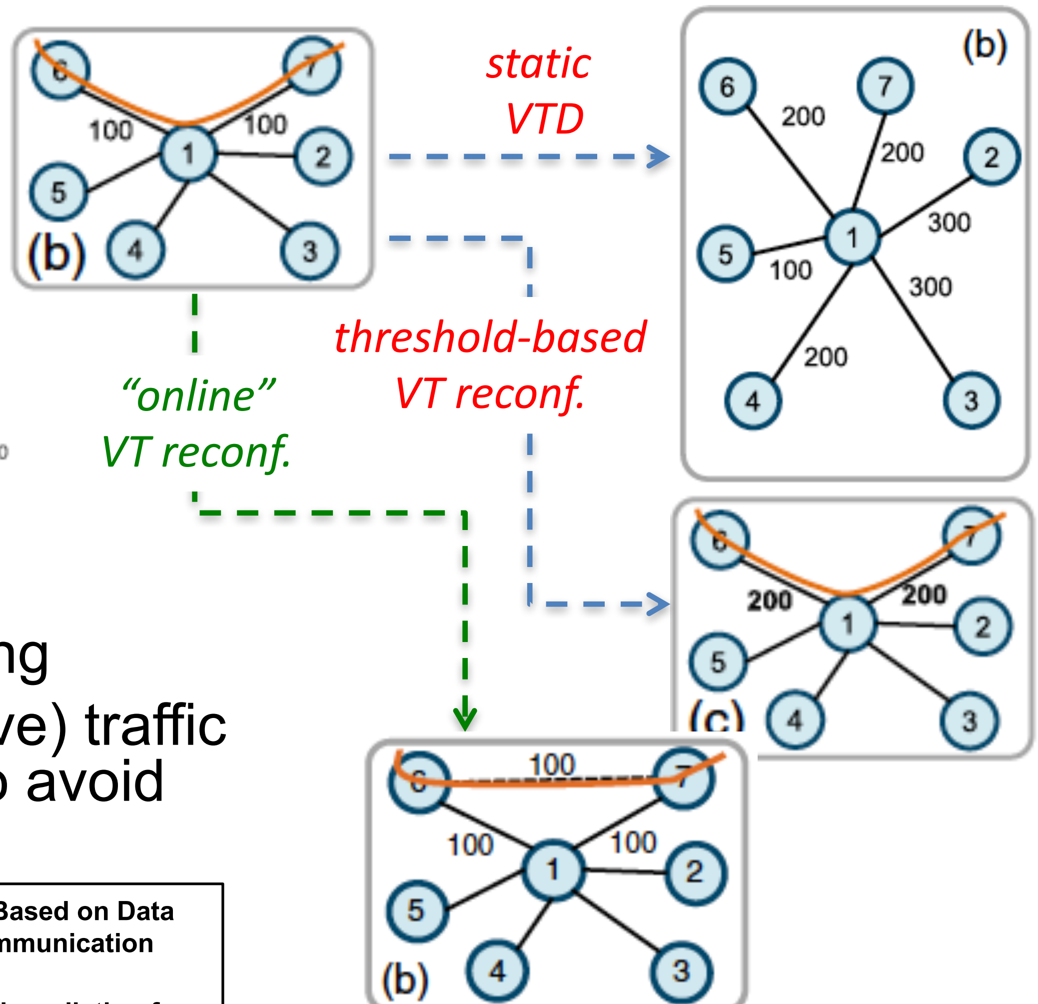
Wang et al., "Nonlinear Decision Boundary Created by a Machine Learning-based Classifier to Mitigate Nonlinear Phase Noise", in ECOC 2015, Sep. 2015

Wang et al., "Nonlinearity Mitigation Using a Machine Learning Detector Based on k-Nearest Neighbors", Photonics Technology Letters, vol. 28 n. 19, Oct. 2016

Network layer
Traffic prediction and virtual topology design

- New services with high spatio-temporal traffic dynamics





- No reconfiguration → peak-traffic dimensioning
- ML leverages *online* (live) traffic monitoring/prediction to avoid overprovisioning

Morales et al., "Virtual Network Topology Adaptability Based on Data Analytics for Traffic Prediction", Journal of Optical Communication and Networking, vol. 9 n. 1, Jan. 2017

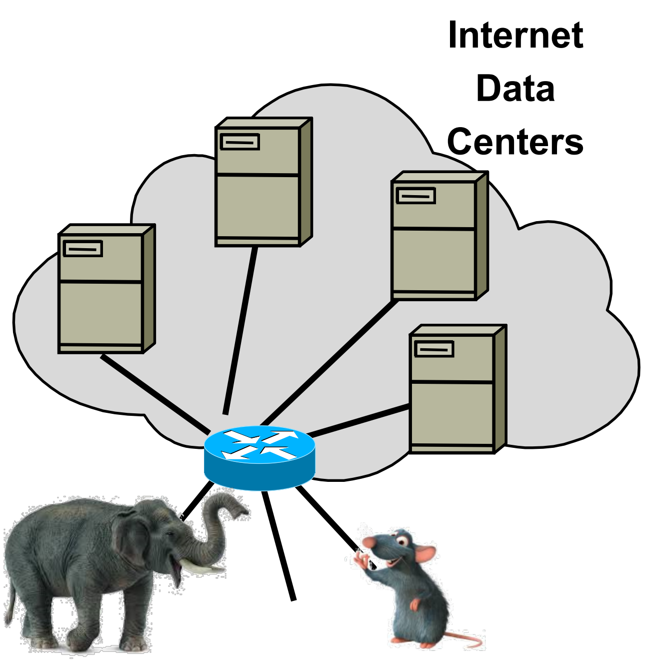
Alvizu et al., "Matheuristic with machine learning-based prediction for software-defined mobile metro-core networks", Journal of Optical Communication and Networking, vol. 9 n. 9, Sep. 2017

Network layer domain
Flow classification

- Traffic flows can be heterogeneous in terms of:
  - protocols (http, ftp, smtp…)
  - services (fixed vs mobile, VoD, data transfer, text messages…)
  - requirements (latency, bandwidth, jitter…)
  - network "customers" (human end-users, companies, sensors, servers…)
    - E.g., "mice" vs "elephant" flows in Data Centers

Distinguish between different flows is crucial for resources (i.e., capacity) allocation, scheduling, SLAs, QoS…

- ML
  - enables traffic features extraction from direct
  - observation of traffic flows
  - allows simultaneous use of heterogeneous features



- **L. Wang, X. Wang, M. Tornatore, K. Joon Kim, S.-M. Kim, D.-U Kim, K.-E. Han, and B. Mukherjee, "Scheduling With Machine-Learning-Based Flow Detection for Packet-Switched Optical Datacenter Networks, to appear JOCN2018**
- **Viljoen et al., "Machine Learning Based Adaptive Flow Classification for Optically Interconnected Data Centers", in ICTON 2016, July 2016**
- **Cao et al., "An accurate traffic classification model based on support vector machines", International Journal on Network Management, 27:e1962, 2017.**

Conclusion

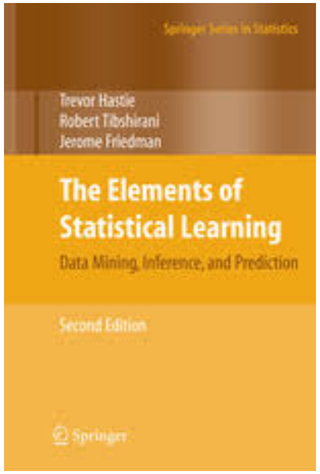
- Automated QoT estimation via machine learning
  - Necessary in a dynamic context
  - Enables margin compression

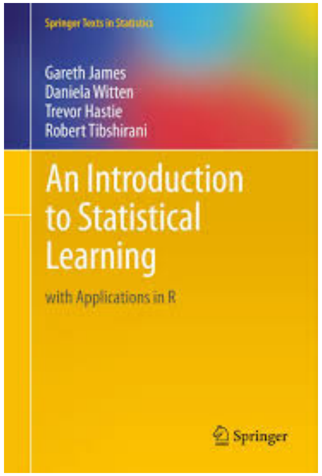
- Automated soft-failure detection and identification
  - Successful testing of identification of failure modes
  - Sampling BER each few seconds led to satisfactory accuracies
  - Identification is more complex than detection (to be confirmed..)

- Definitely not a fast learning curve if you do not simply want to use machine learning as a black box

Some material

- Books (general refs. for ML):
  - T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning”, Ed. Springer
  - G. James, D. Witten, T. Hastie, R. Tibshirani, “An Introduction to Statistical Learning with Applications in R”, Ed. Springer





- Prof. Andrew Ng lectures (Stanford University)
- … Google it!

Some publications (1)

Surveys

- F. Musumeci et al., “A Survey on Application of Machine Learning Techniques in Optical Networks”, Submitted to IEEE communication surveys and tutorials
- Javier Mata, et a., Artificial intelligence (AI) methods in optical networks: A comprehensive survey, Optical Switching and Networking, Volume 28, 2018, pp. 43-57

Some Motivations

- Y. Pointurier, "Design of low-margin optical networks," in *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 1, pp. A9-A17, Jan. 2017. doi: 10.1364/JOCN.9.0000A9

QoT estimation

- Barletta et al., “QoT Estimation for Unestablished Lighpaths using Machine Learning”, in OFC 2017 Conference, Mar. 2017
- De Miguel *et al.*, “Cognitive Dynamic Optical Networks”, *Journal of Optical Communication and Networking*, vol. 5, n. 10, Oct. 2013
- Thrane *et al.*, “Machine Learning Techniques for Optical Performance Monitoring From Directly Detected PDM-QAM Signals”, *Journal of Lightwave Technology*, vol. 35, n. 4, Feb. 2017
- Caballero *et al.*, “Experimental demonstration of a cognitive quality of transmission estimator for optical communication systems”, *Optics Express*, vol. 20, n. 26, Dec. 2012
- Jimenez *et al.*, “A Cognitive Quality of Transmission Estimator for Core Optical Networks”, *Journal of Lightwave Technology*, vol. 31, n. 6, Mar. 2013
- Angelou *et al.*, “Optimized Monitor Placement for Accurate QoT Assessment in Core Optical Networks”, *Journal of Optical Communication and Networking*, vol. 4, n. 1, Jan. 2012

Some publications (2)

Failure recovery

- S. Shahkarami, F. Musumeci, F. Cugini, M. Tornatore, \Machine-Learning-Based Soft-Failure Detection and Identi cation in Optical Networks,"in Proceedings, OFC 2018, San Diego (CA), Usa, Mar. 11-15, 2017
- A. Vela *et al*., "Soft Failure Localization during Commissioning Testing and Lightpath Operation", *Journal of Optical Communication and Networking*, vol. 10 n. 1, Jan. 2018
- A. Vela *et al*., "BER degradation Detection and Failure Identification in Elastic Optical Networks", in Journal of Lightwave Technology, vol. 35, no. 21, pp. 4595-4604, Nov.1, 1 2017

Projects

- EU ORCHESTRA and CHRON projects

Thank You!

..and thanks to them!

POLITECNICO MILANO 1863

Achille Pattavina
Francesco Musumeci
Shahin Shahkarami
Luca Barletta

Biswanath Mukherjee, Yu Wu
Lin Wang,Sabidur Rehman,

Filppo Cugini (CNIT)
Cristina Rottondi and Alessandro Giusti (Dalle Molle Inst.)

European Commission

Horizon 2020
European Union funding
for Research & Innovation