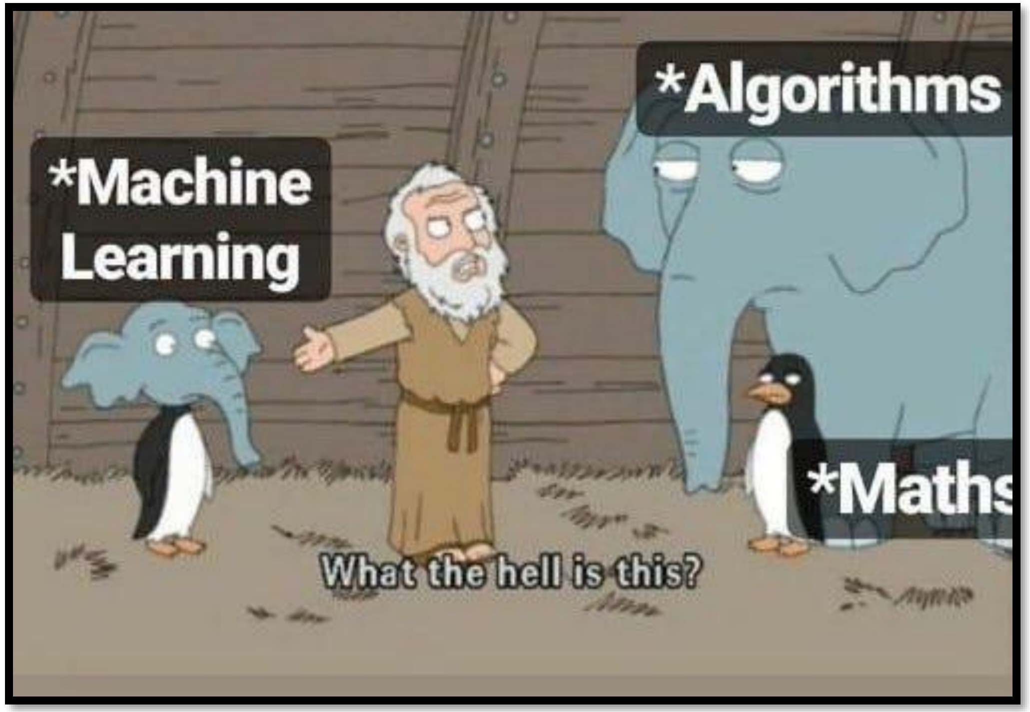
OTIMIZANDO A DETECÇÃO DE CIBERATAQUES COM ENSEMBLE LEARNING

Thiago José Lucas, Ph.D.

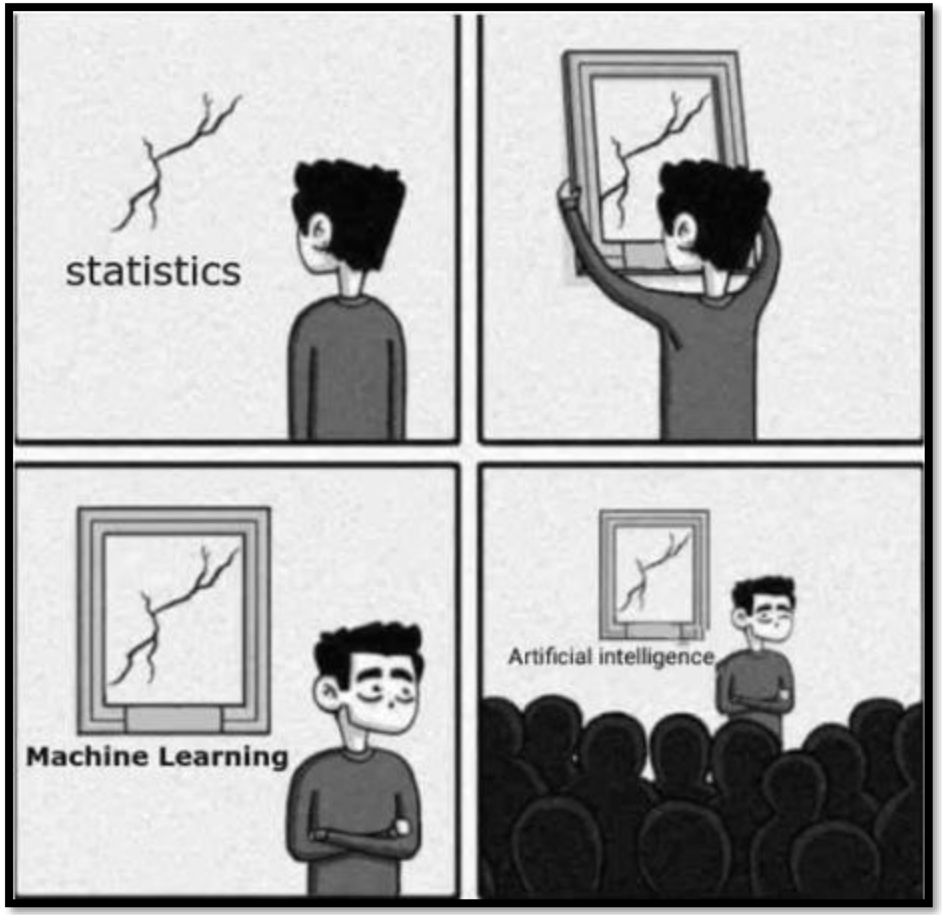
O QUE É "APRENDIZAGEM DE MÁQUINA"?



Aplicações tradicionais

- Classificação
- Regressão
- Clustering

O QUE É "APRENDIZAGEM DE MÁQUINA"?



Como a máquina aprende?

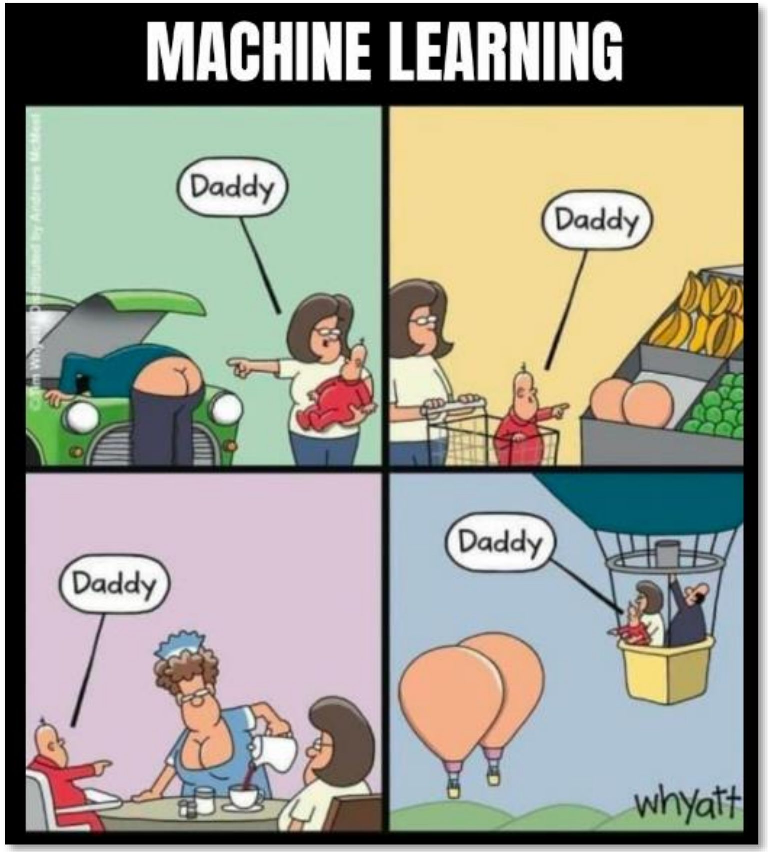
- Com supervisão
- Sem supervisão
- Semi-supervisão, reforço, etc

QUAL A IMPORTÂNCIA NO BLUE TEAM?

Evolução

- Mudança brusca do blue team raiz ☺
- A complexidade dos ataques recentes
- Seus mecanismos de defesa estão adaptados aos 0-days?
- O que não é benígno, é malicioso!

COMO USAR EM BENEFÍCIO DO BLUE TEAM?



Detecção de Intrusão

- Reconhecimento de ataques

COMO USAR EM BENEFÍCIO DO BLUE TEAM?

Redução dos erros de classificação

- IDSs tradicionais erram mais (muito mais!)
- É possível errar menos (quase nada)
    - Datasets confiáveis
    - Pré-processamento adequado
    - Algoritmos robustos e otimizados

COMO USAR EM BENEFÍCIO DO BLUE TEAM?

- Evolução
- Relação "IA/IDS"

Firewall clássico → Firewall stateful → Firewall *layer* 7 → Sistemas de Detecção de Intrusão

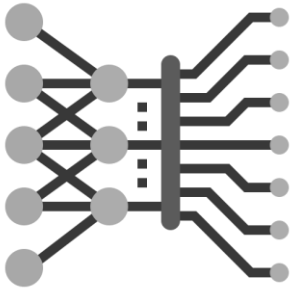
↓

Host-based IDS
Network-based IDS

↓

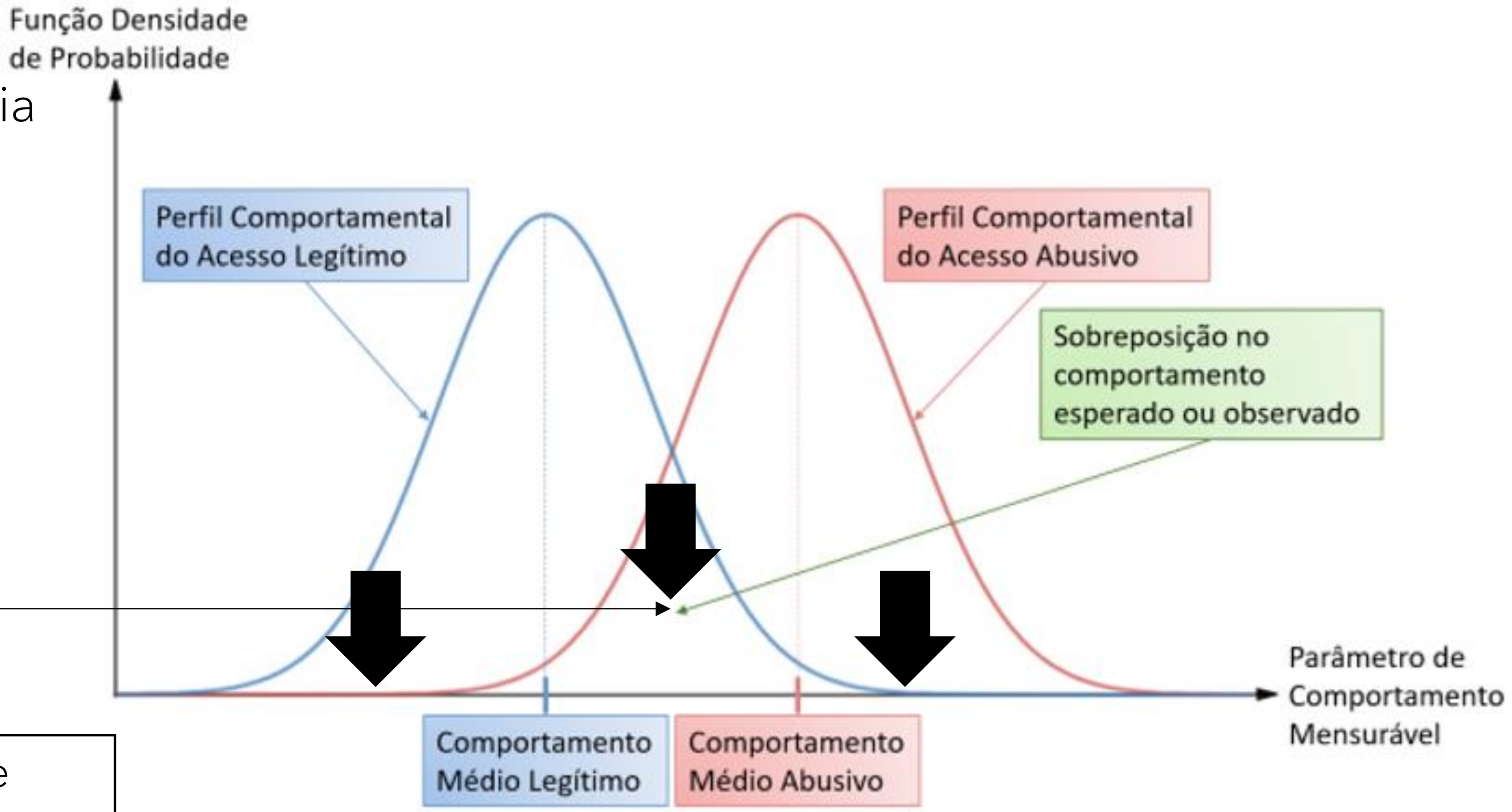
Assinatura
Anomalia

- Métodos matemático/estatísticos
  - Reconhecimento de padrões complexos
    - Modelos robustos



SISTEMAS DE DETECÇÃO DE INTRUSÃO

- HIDS e NIDS
- Assinatura e anomalia

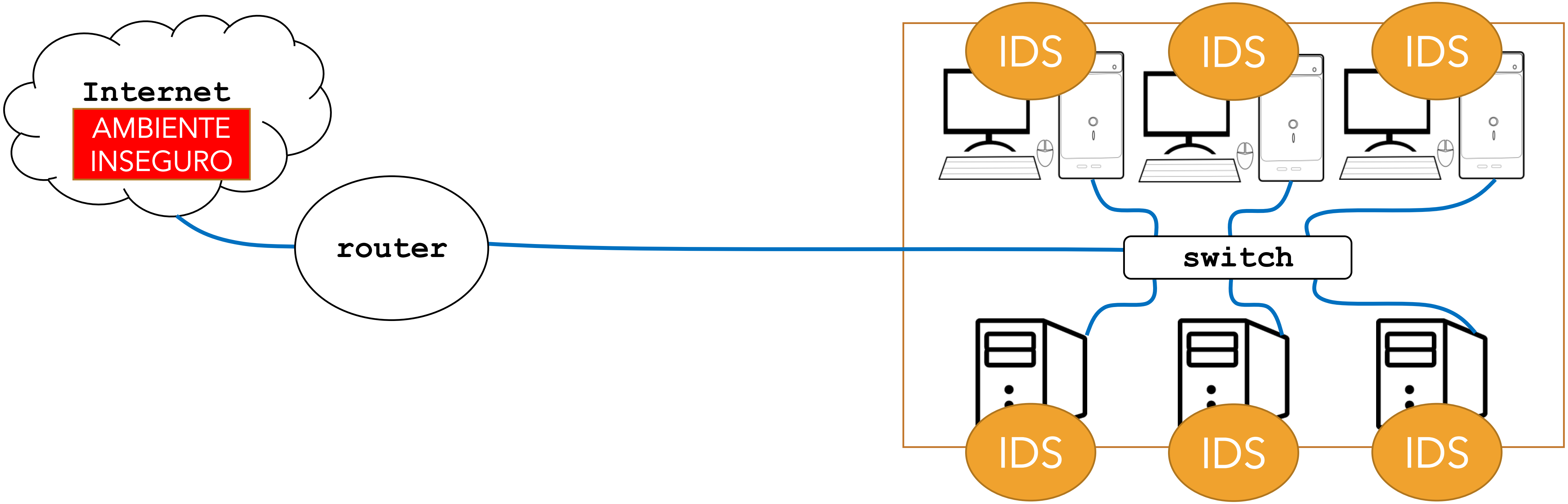


- Diminuição dos erros de classificação (FN / FP)
- Consequente incremento de acurácia / performance

SISTEMAS DE DETECÇÃO DE INTRUSÃO

- HIDS e NIDS
- Assinatura e anomalia

Host-based IDS

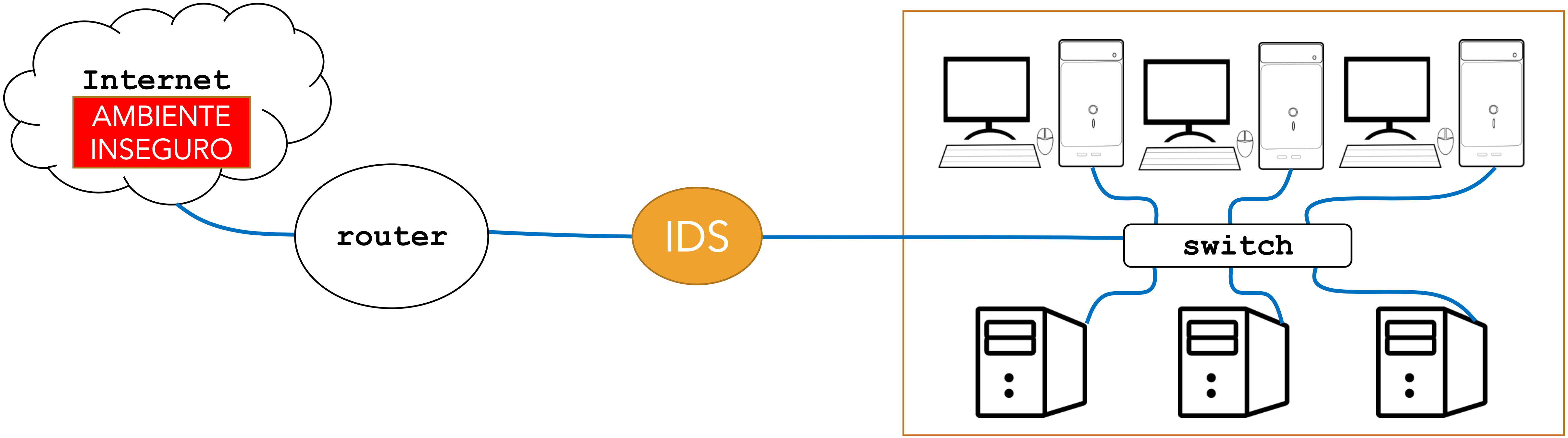


SISTEMAS DE DETECÇÃO DE INTRUSÃO

- HIDS e NIDS
- Assinatura e anomalia

Network-based IDS

- Bridge
- Espelhamento

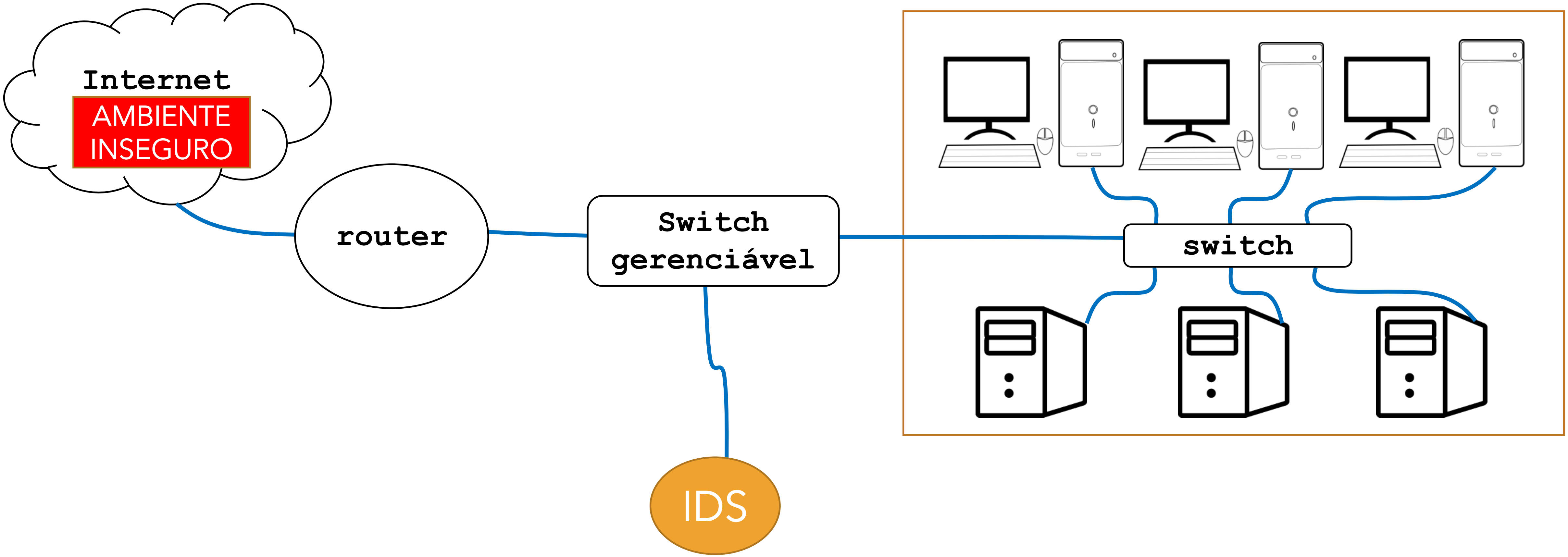


SISTEMAS DE DETECÇÃO DE INTRUSÃO

- HIDS e NIDS
- Assinatura e anomalia

Network-based IDS

- Bridge
- Espelhamento



E COMO SE FAZ ISSO?

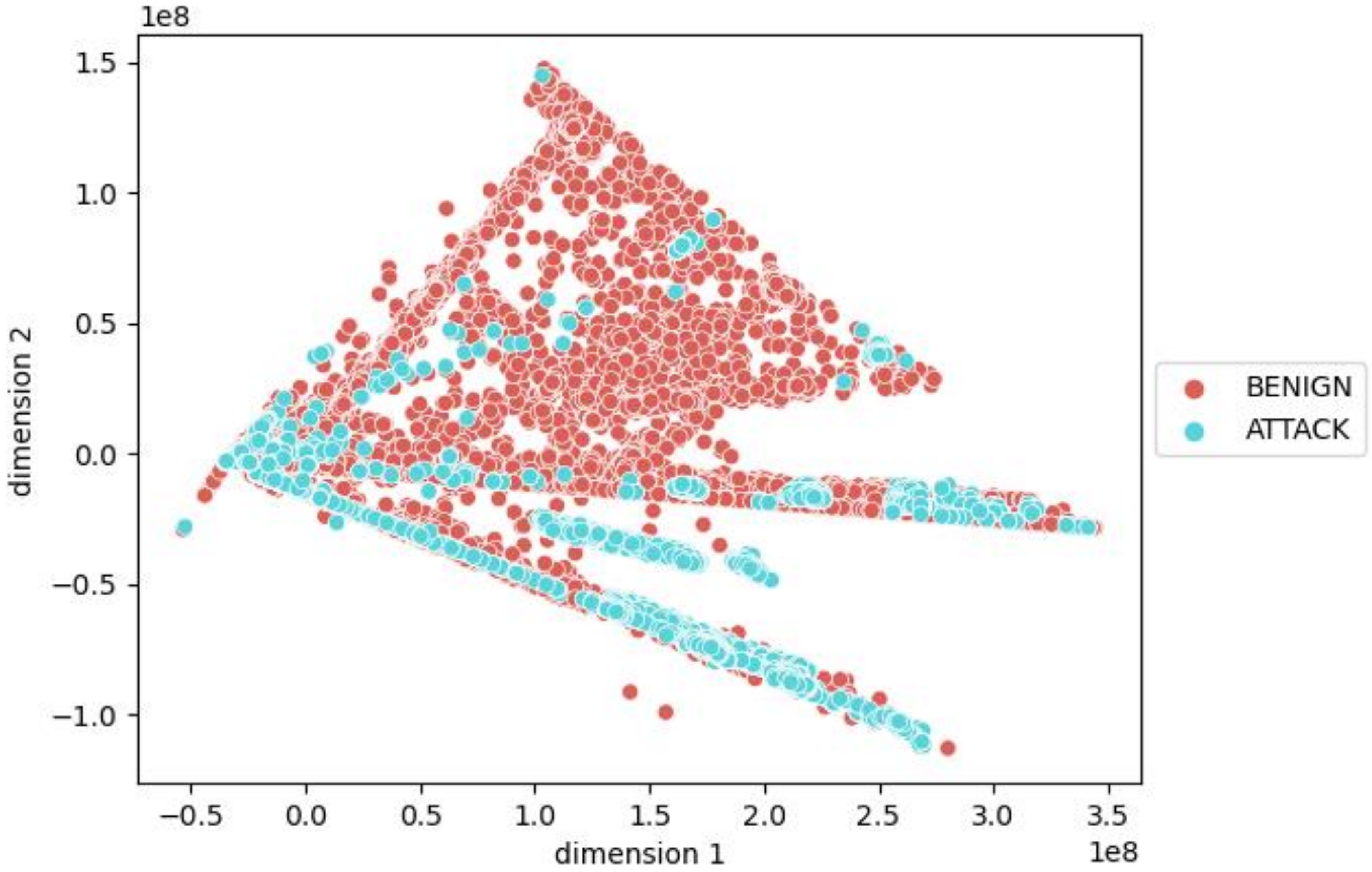
DATASET

Mostrar um dataset aleatório

DATASET

Category		Total	Total(-rows with lack info)	Training	Test
BENIGN	BENIGN	2273097	2271320	20000	20000
DOS	DDoS	128027	128025	2700	3300
	DoS slowloris	5796	5796	1350	1650
	DoS Slowhttptest	5499	5499	2171	1169
	DoS Hulk	231073	230124	4500	5500
	DoS GoldenEye	10293	10293	1300	700
	Heartbleed	11	11	5	5
PortScan	PortScan	158930	158804	3808	4192
Bot	Bot	1966	1956	936	624
Brute-Force	FTP-Patator	7938	7935	900	1100
	SSH-Patator	5897	5897	900	1100
Web Attack	Web Attack-Brute Force	1507	1507	910	490
	Web Attack-XSS	652	652	480	160
	Web Attack-SQL Injection	21	21	16	4
Infiltration	Infiltration	36	36	24	6
Total Attack		471454	470365	20000	20000
Total		2830743	2827876	40000	40000

DATASET



1e8
1.5
1.0
0.5
0.0
−0.5
−1.0
dimension 2
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
dimension 1
1e8
BENIGN
ATTACK

TREINAMENTO

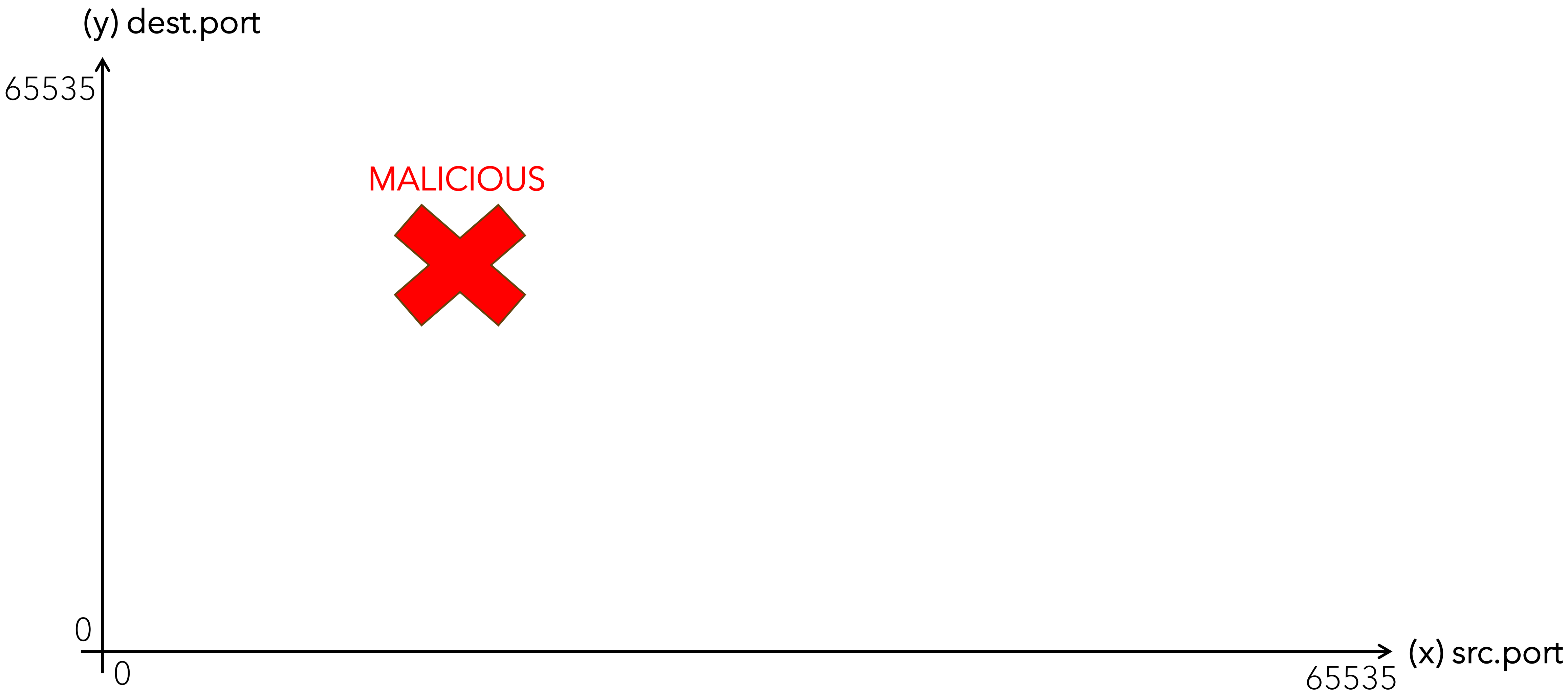
(x) src.port

TREINAMENTO

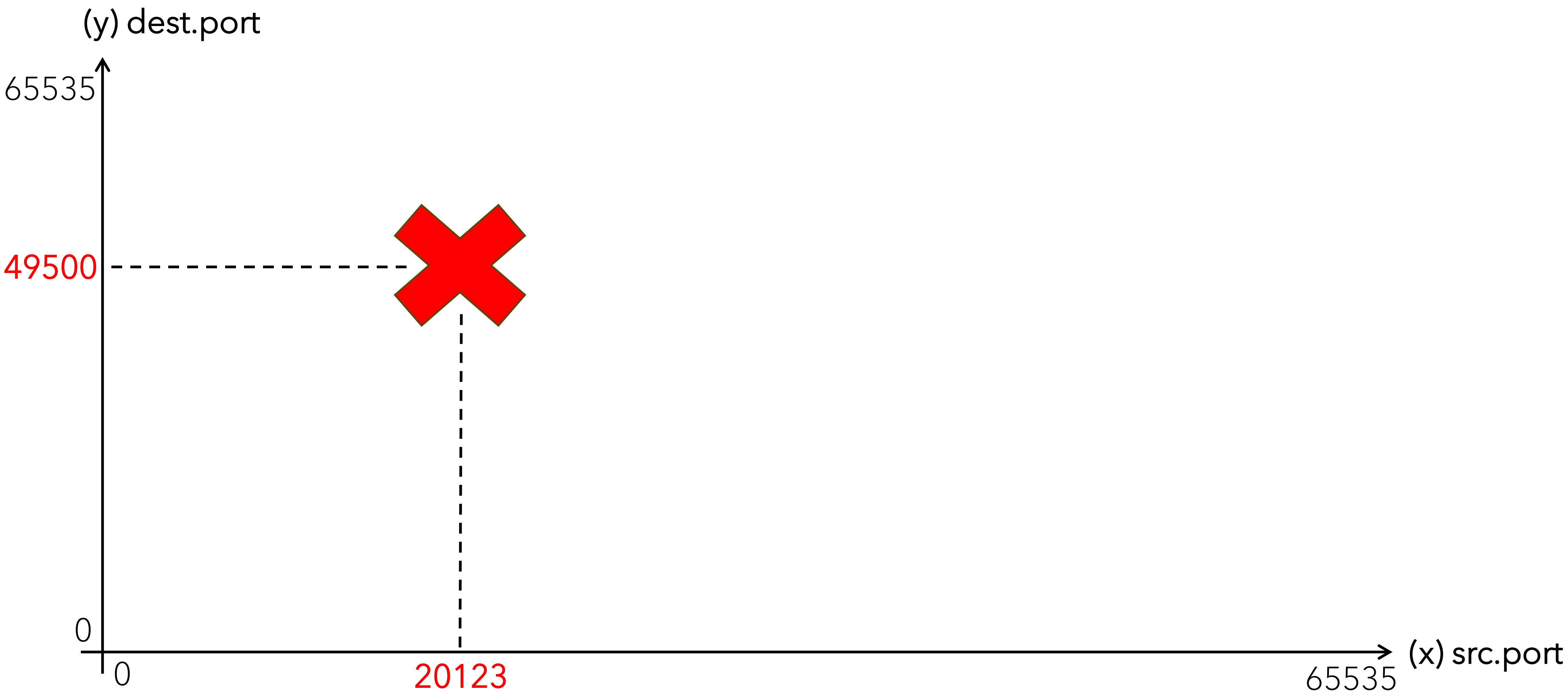
TREINAMENTO

TREINAMENTO

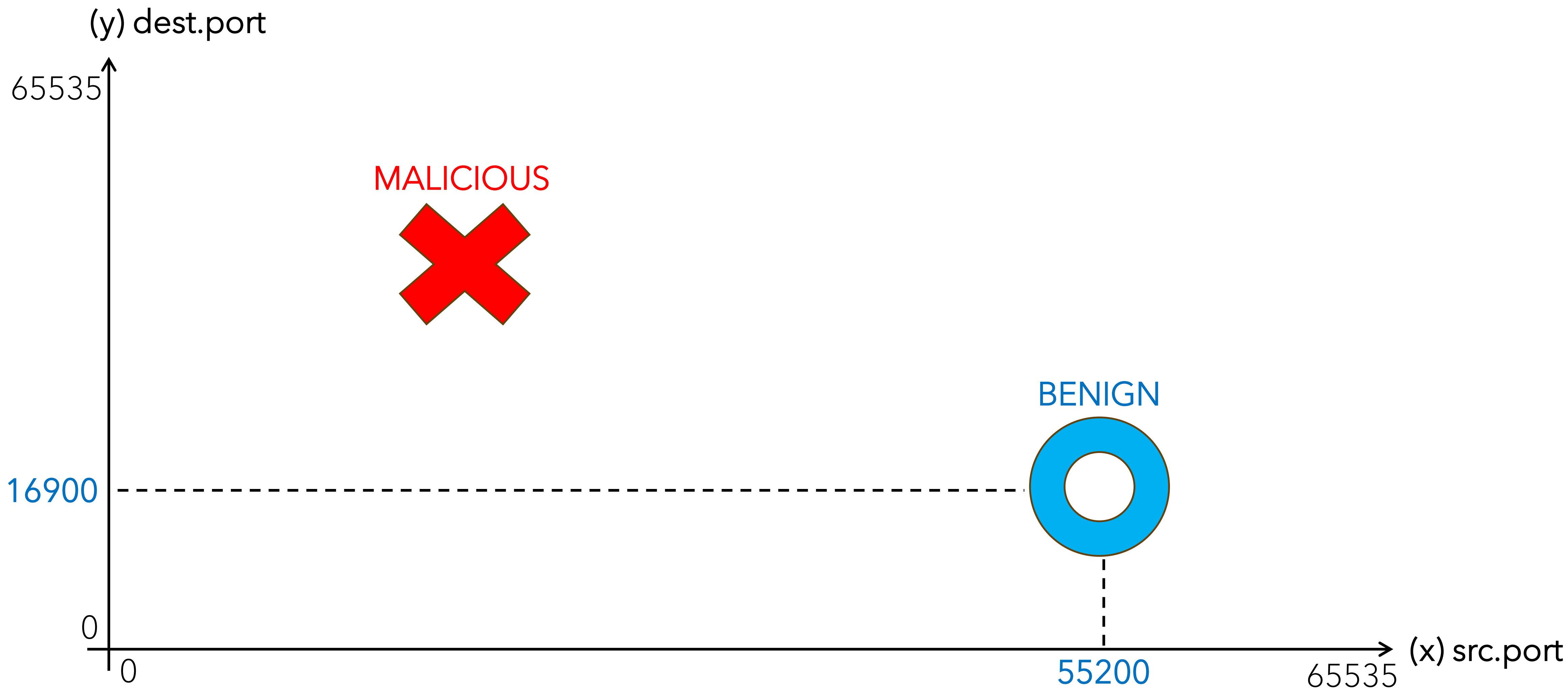
TREINAMENTO



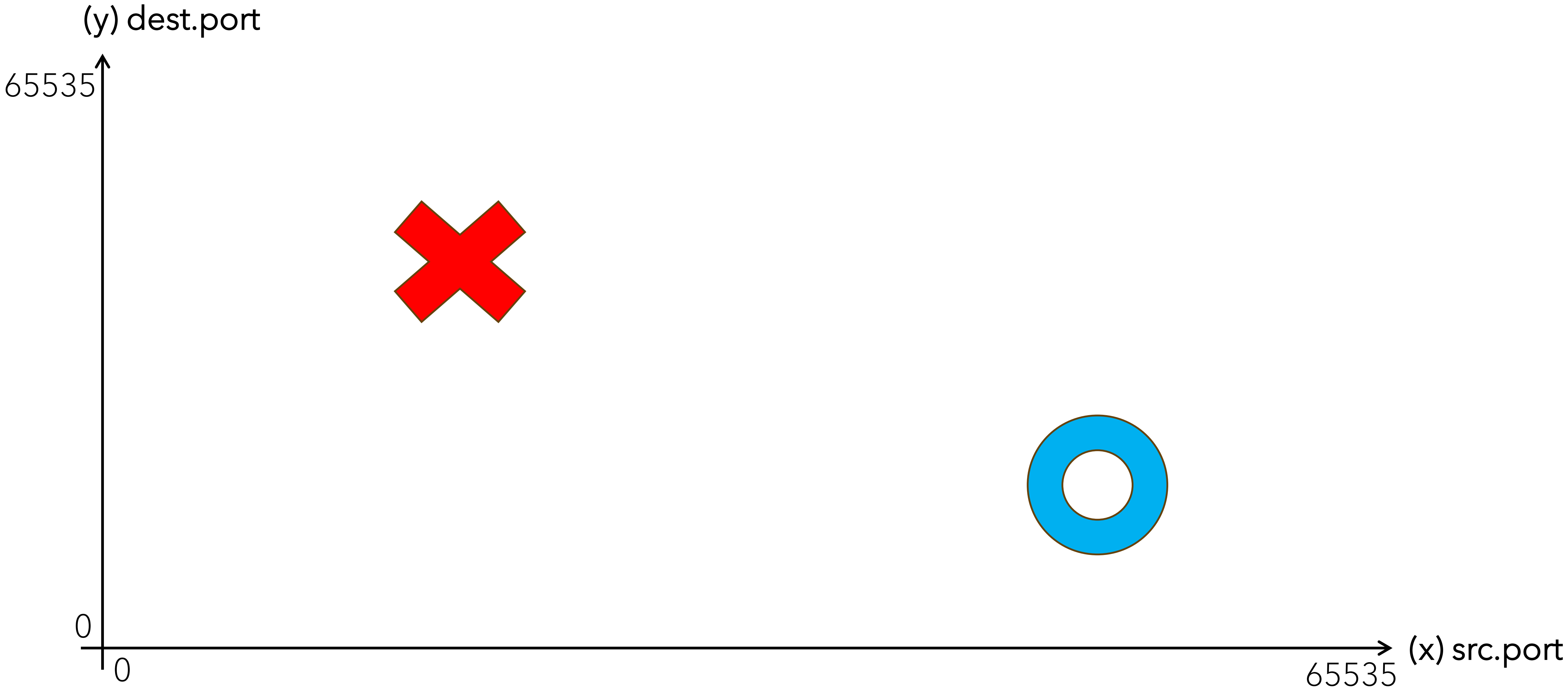
TREINAMENTO



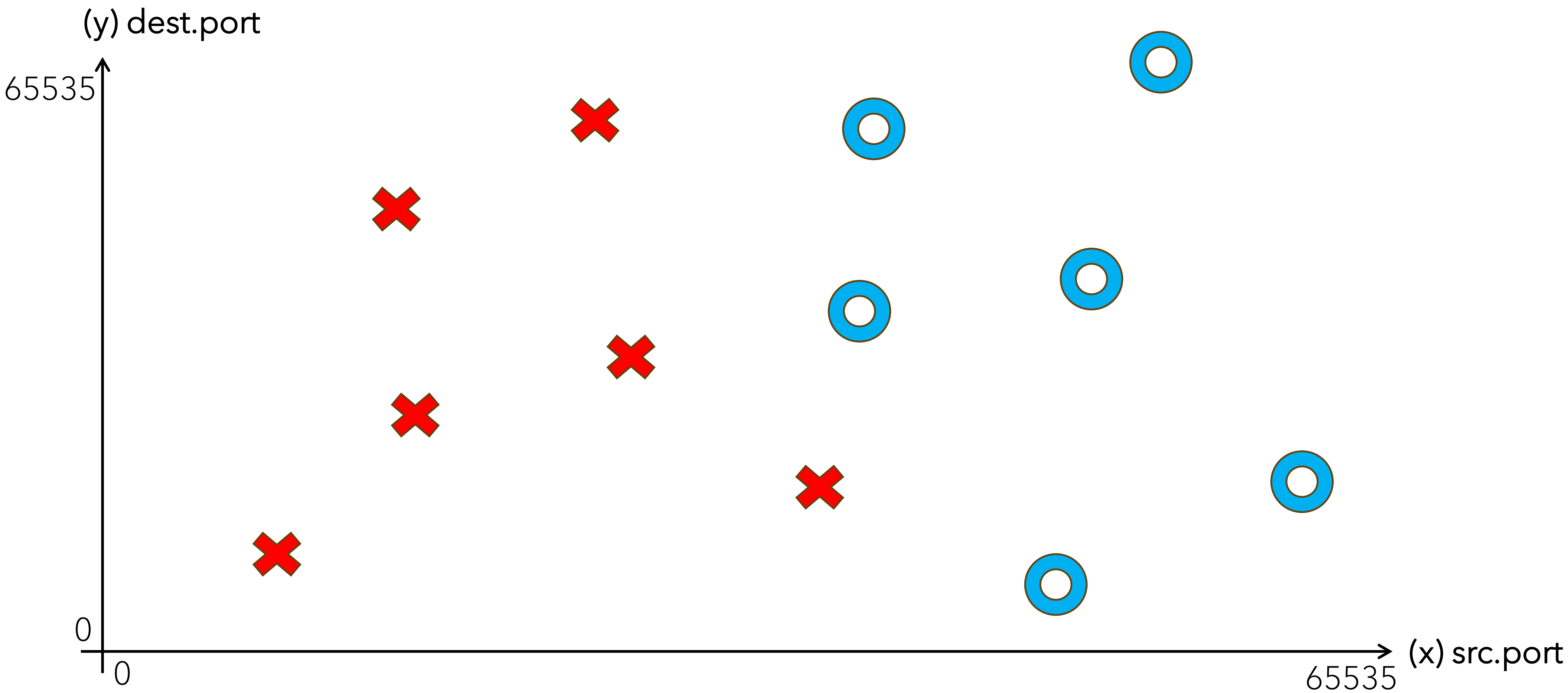
TREINAMENTO



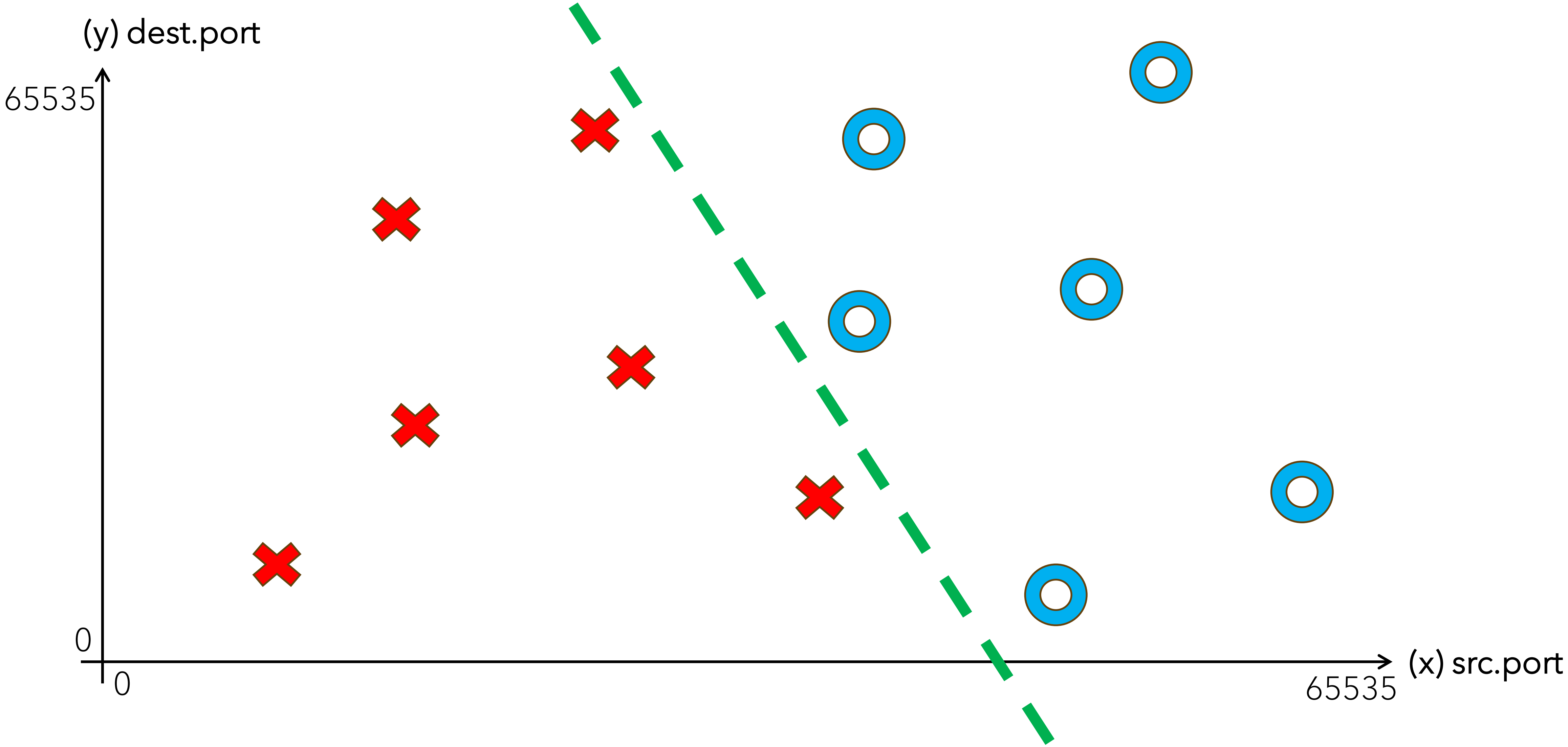
TREINAMENTO



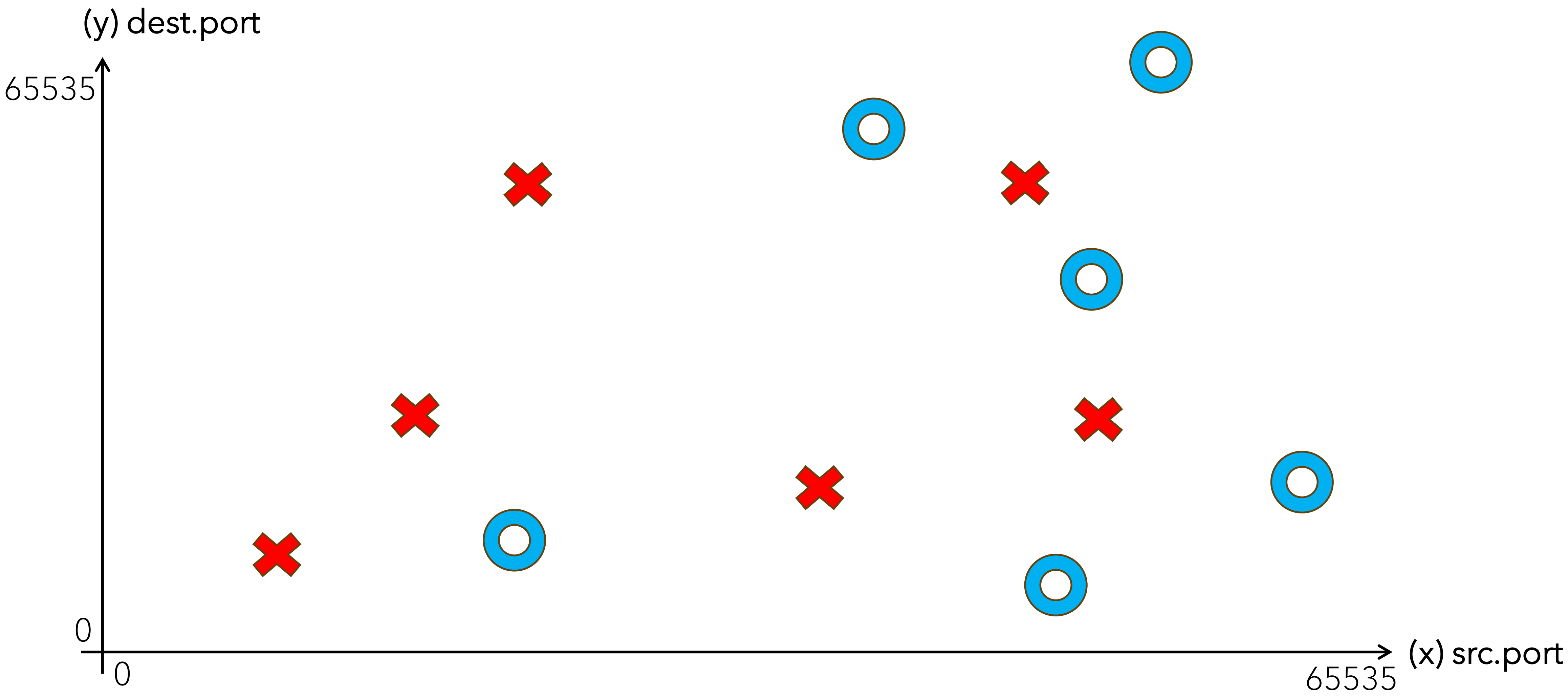
TREINAMENTO – PROBLEMA SIMPLES, SOLUÇÃO LINEARMENTE SEPARÁVEL



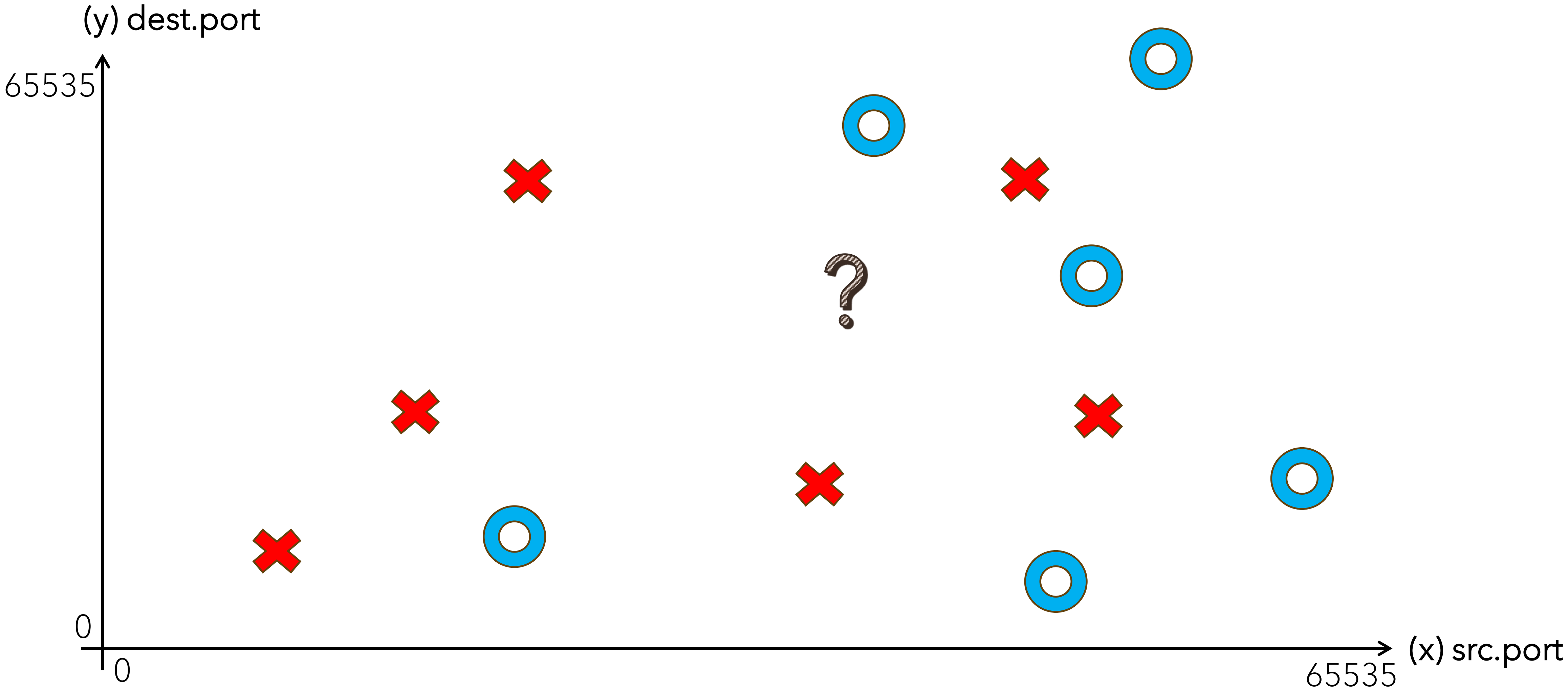
TREINAMENTO – PROBLEMA SIMPLES, SOLUÇÃO LINEARMENTE SEPARÁVEL



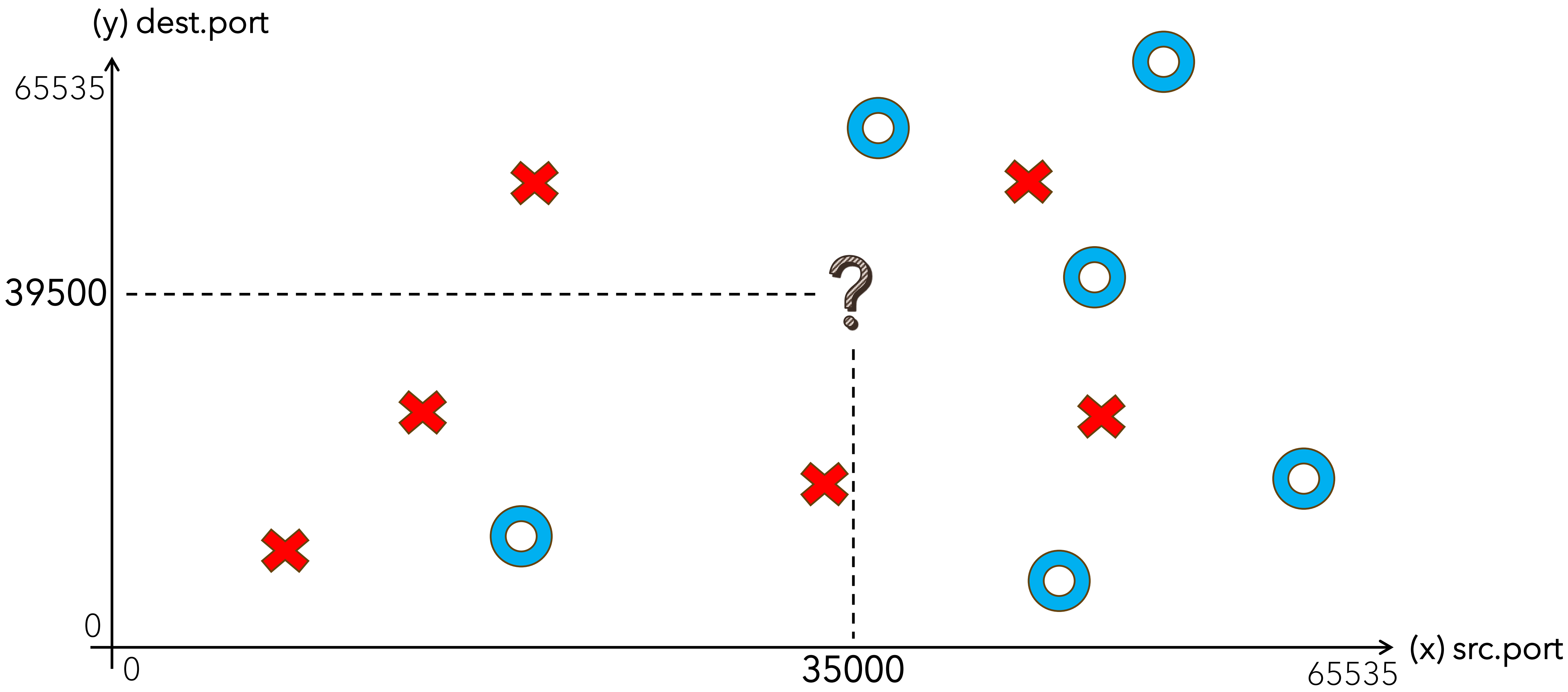
TREINAMENTO – PROBLEMAS COMPLEXOS



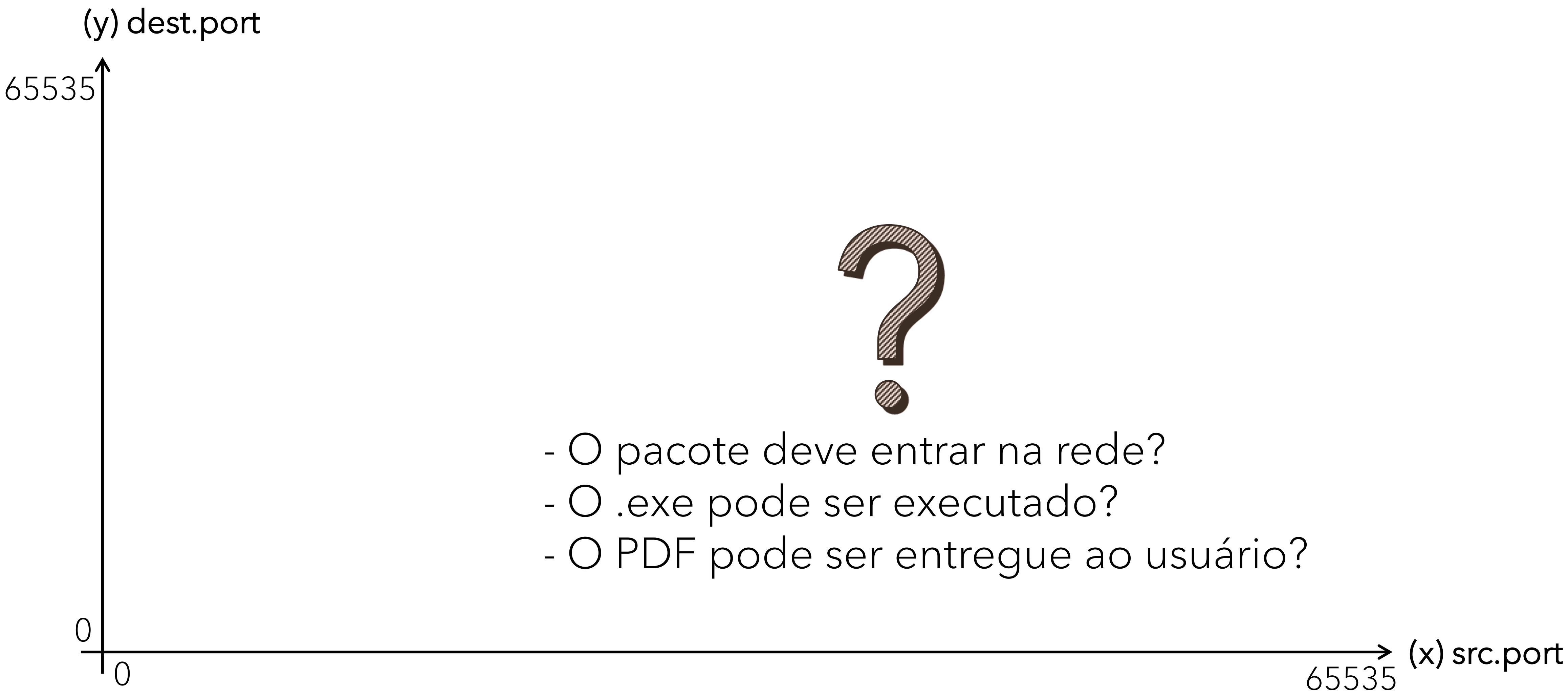
TREINAMENTO – PROBLEMAS COMPLEXOS



TREINAMENTO – PROBLEMAS COMPLEXOS

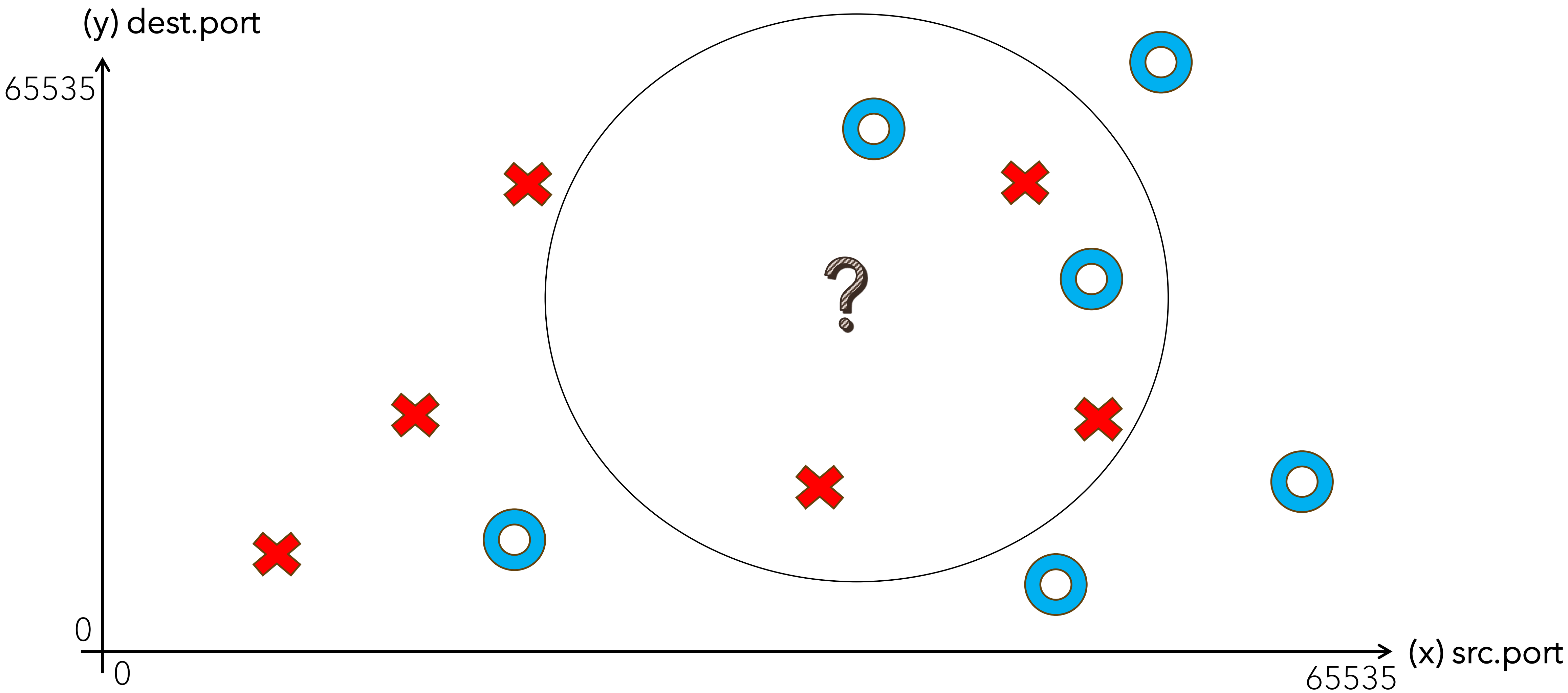


TREINAMENTO – PROBLEMAS COMPLEXOS

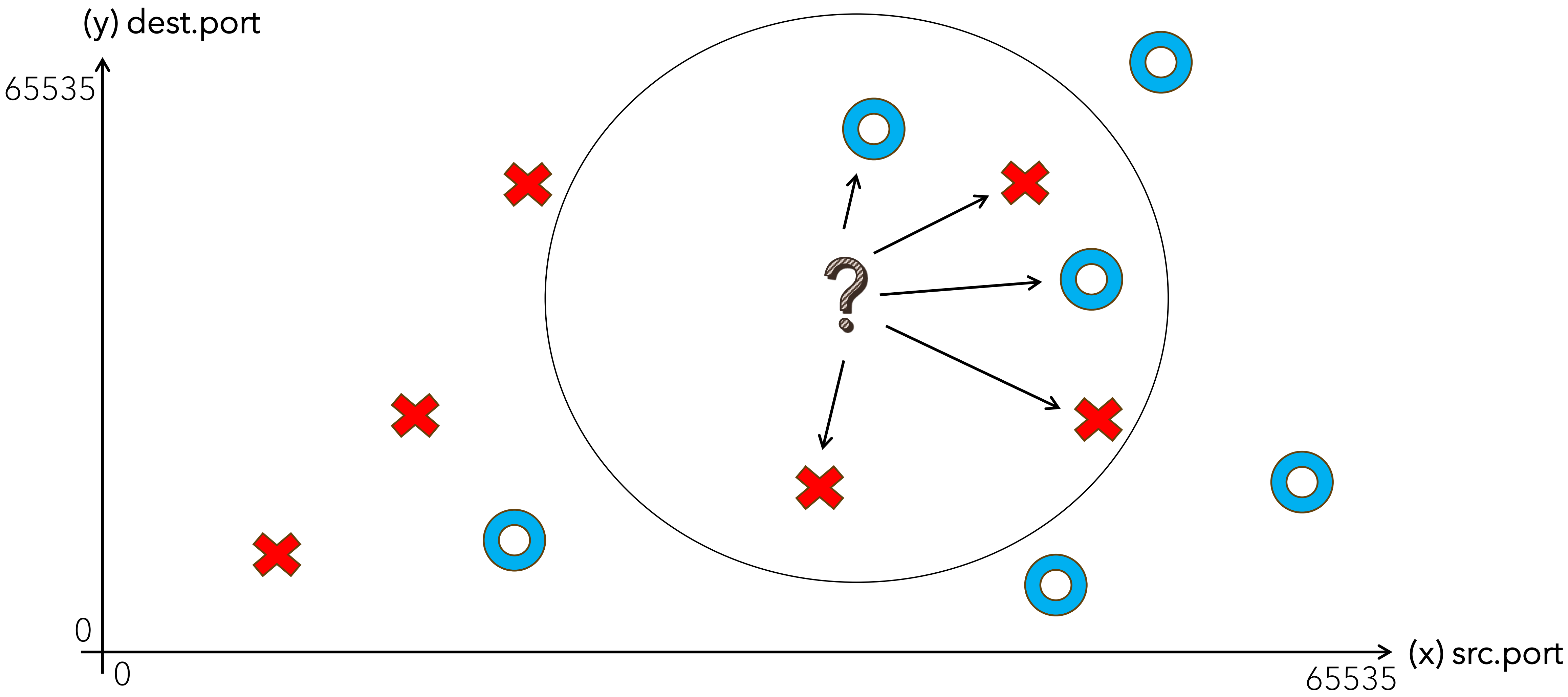


- O pacote deve entrar na rede?
- O .exe pode ser executado?
- O PDF pode ser entregue ao usuário?

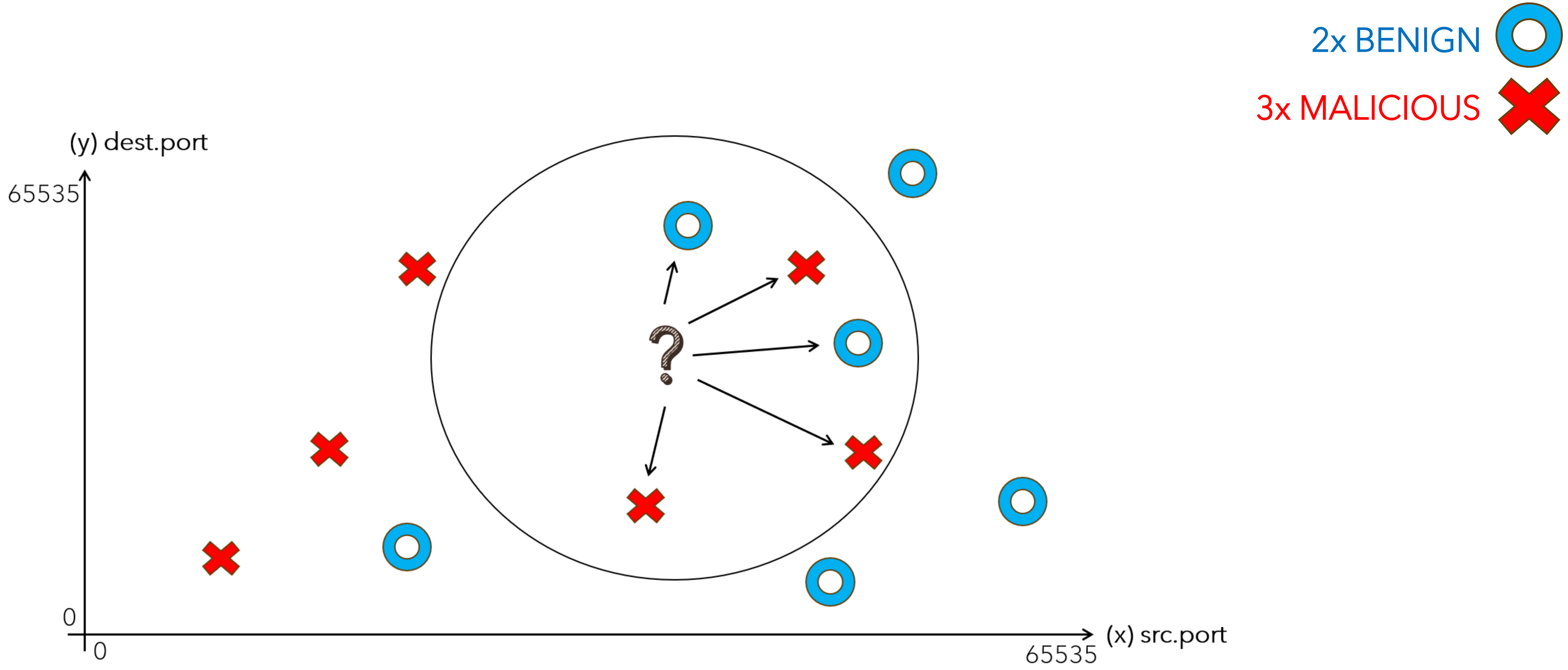
TREINAMENTO – PROBLEMAS COMPLEXOS



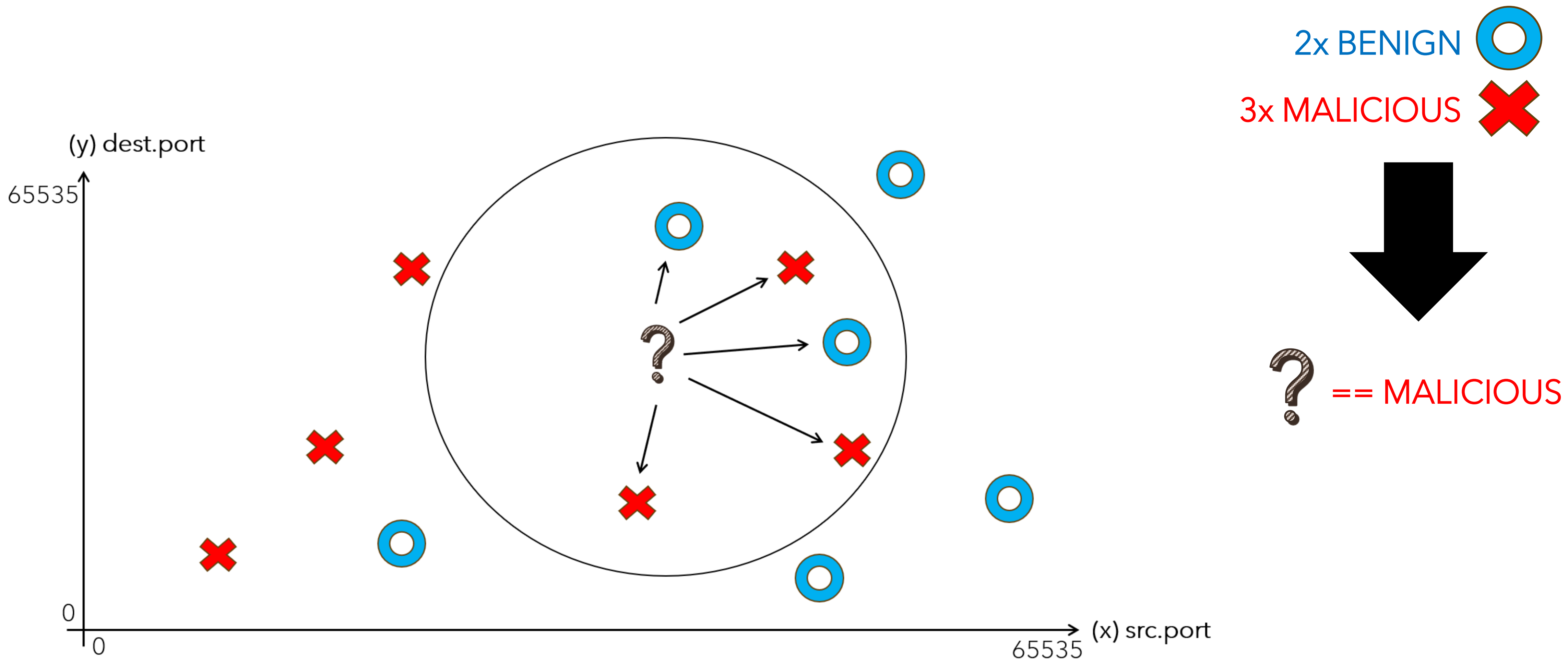
TREINAMENTO – PROBLEMAS COMPLEXOS



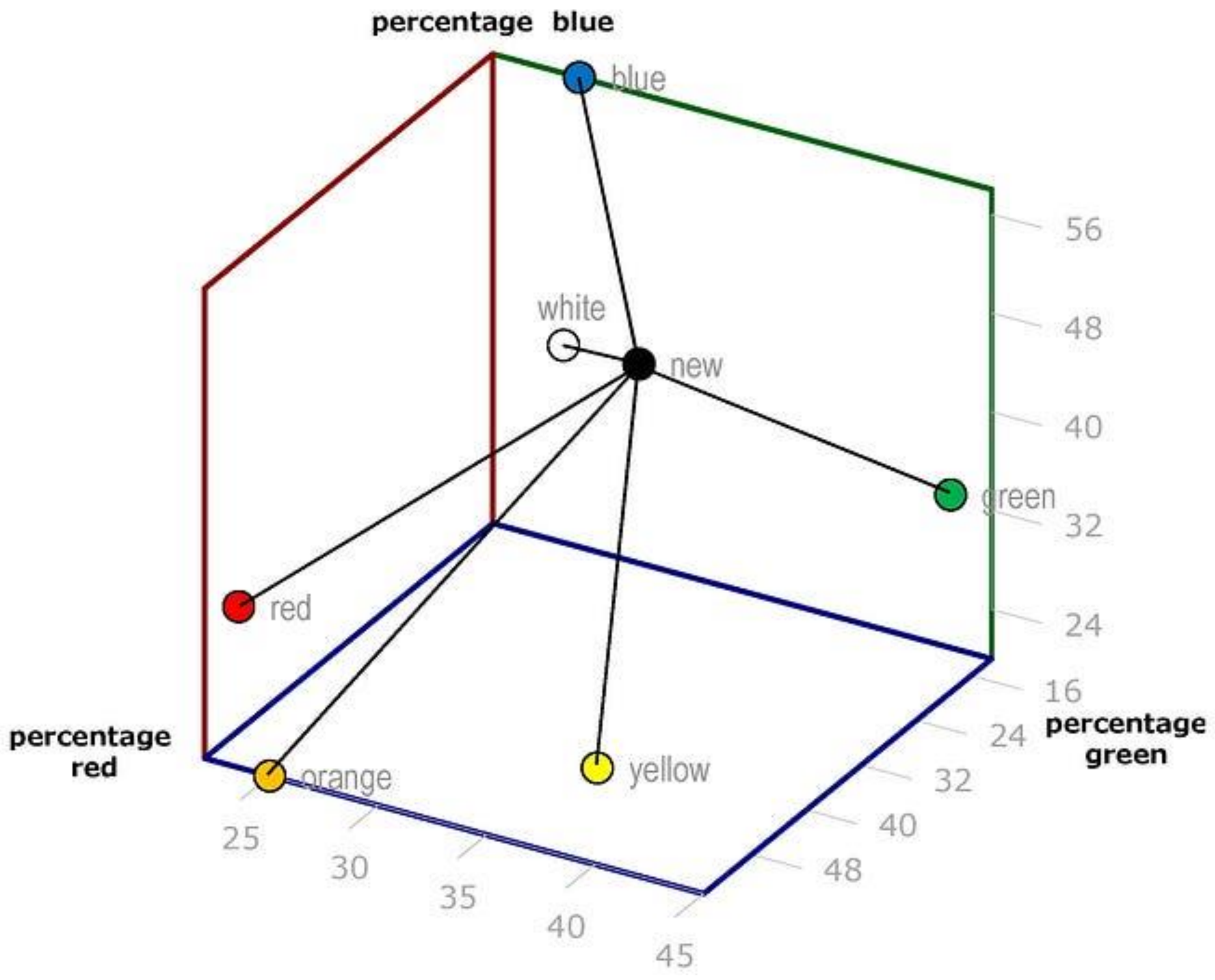
TREINAMENTO – PROBLEMAS COMPLEXOS



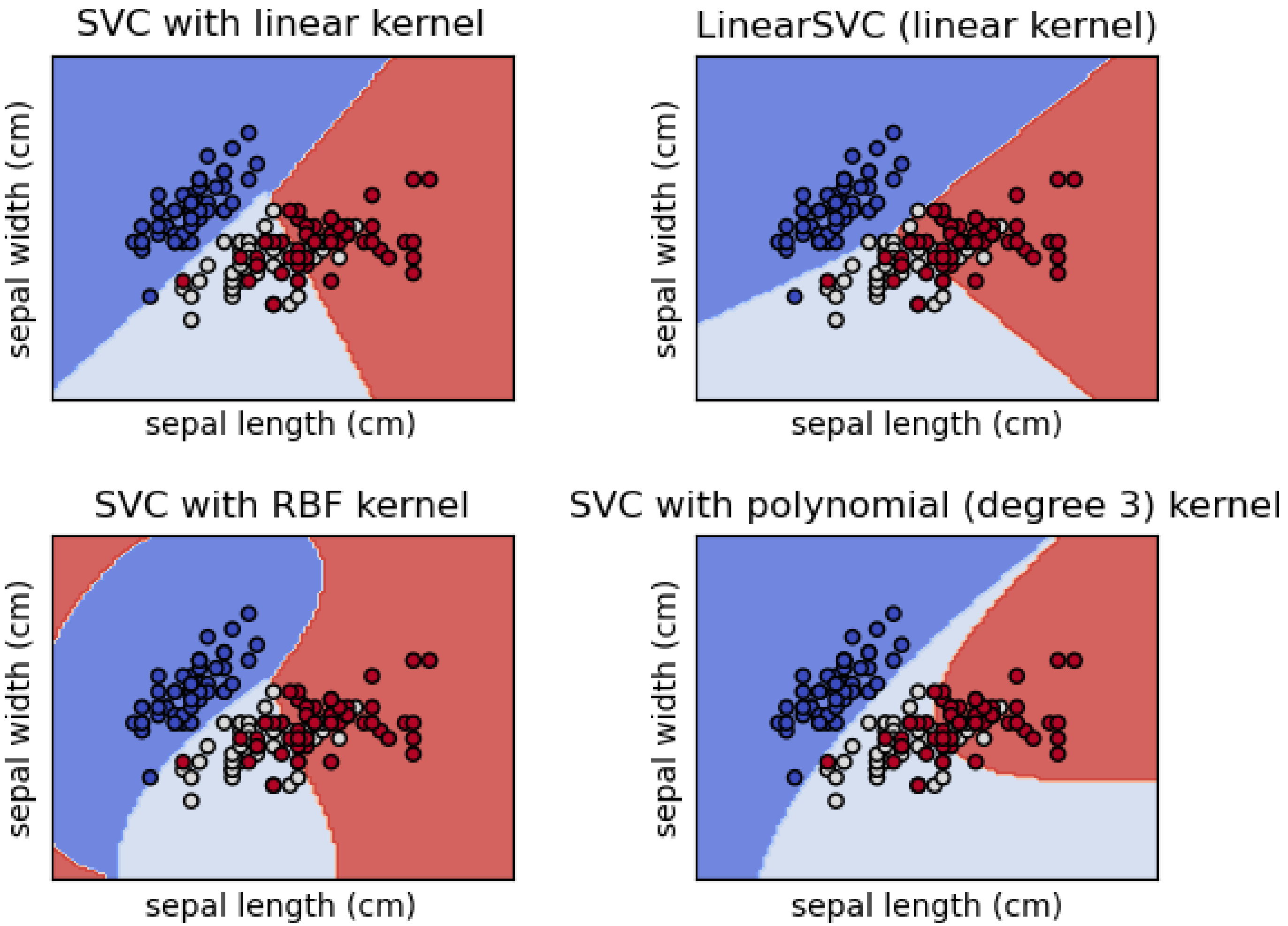
TREINAMENTO – PROBLEMAS COMPLEXOS



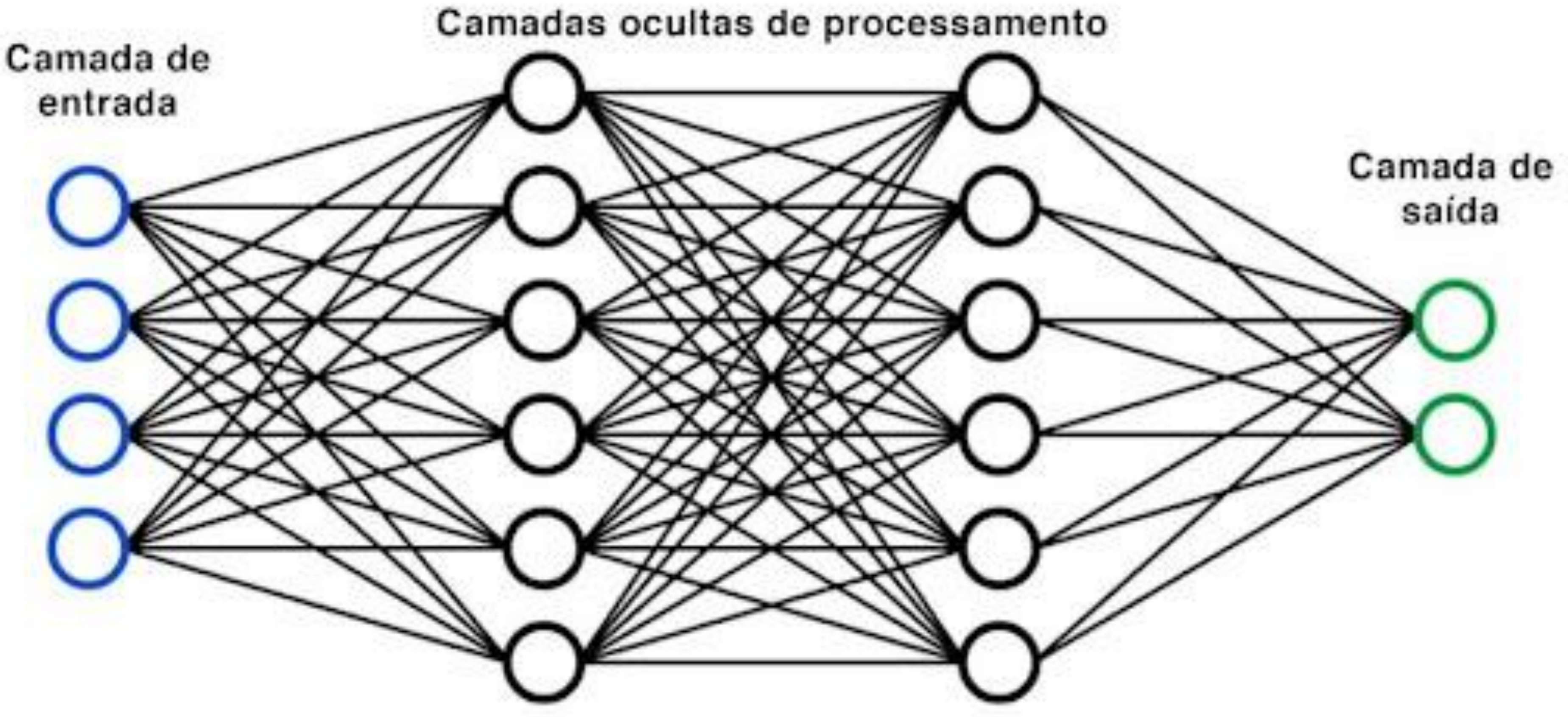
TREINAMENTO – PROBLEMAS COMPLEXOS



TREINAMENTO – PROBLEMAS COMPLEXOS

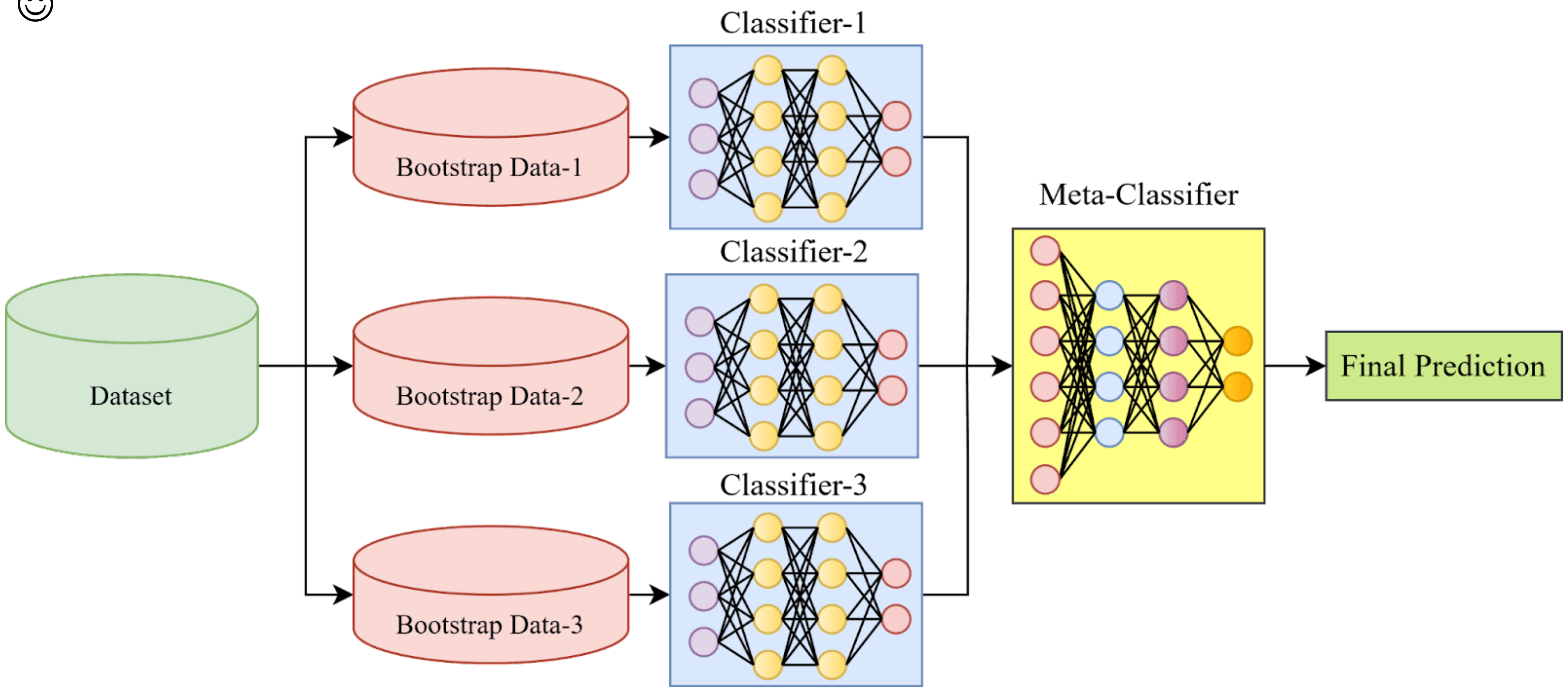


TREINAMENTO – PROBLEMAS COMPLEXOS

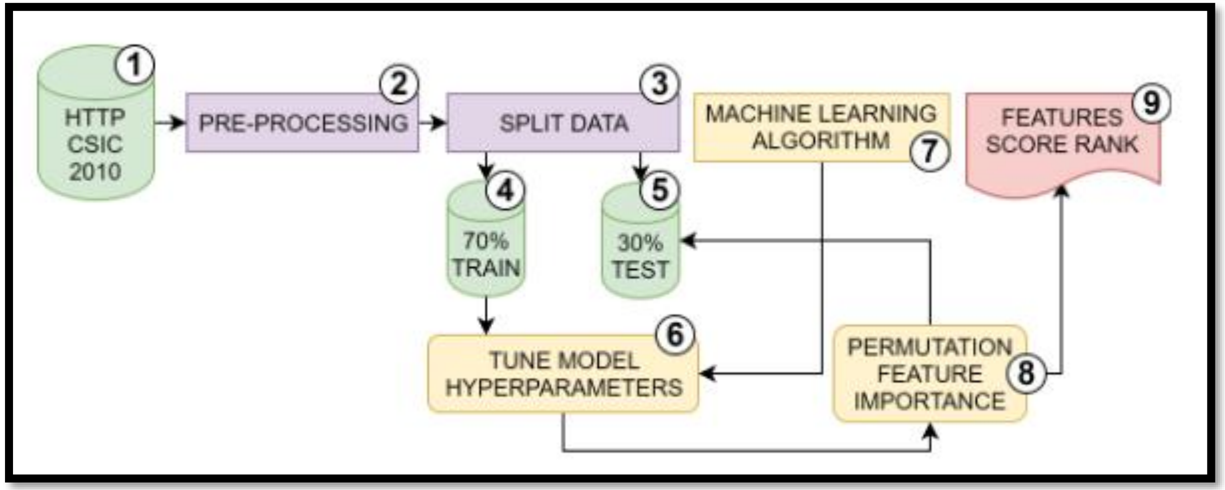


TREINAMENTO – PROBLEMAS + COMPLEXOS AINDA ☺

- Ensemble Learning
- Custo computacional ☹
- Ensemble Pruning ☺



ALGUNS RESULTADOS INTERESSANTES...



Algorithm	Accuracy		Precision		Recall		F1-Score	
SVM	0.95	**0.98**	0.94	**0.97**	0.92	**0.97**	0.93	**0.97**
BPM	0.90	**0.97**	0.89	**0.98**	0.87	**0.94**	0.88	**0.96**
AP	0.83	**0.96**	0.8	**0.96**	0.79	**0.95**	0.78	**0.95**
NN	0.84	**0.99**	0.83	**0.99**	0.82	**0.98**	0.79	**0.99**
DF	0.66	**0.77**	0.68	**0.90**	**0.69**	0.56	0.64	**0.69**
DJ	0.62	**0.70**	0.63	**0.88**	**0.60**	0.43	**0.62**	0.57
BDT	0.64	**0.88**	0.65	**0.94**	0.68	**0.78**	0.65	**0.85**
LR	**0.97**	**0.97**	0.92	**0.97**	0.95	**0.96**	0.96	**0.97**
	mRMR	PFI	mRMR	PFI	mRMR	PFI	mRMR	PFI

ALGUNS RESULTADOS INTERESSANTES...

SBRC 2021

XXXIX Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos

Uberlândia - MG, de 16 a 20 de agosto de 2021

Tabela 4. Compilado com todas as taxas de erro individuais e com o *Stacking* escolhido pela aplicação da combinação proposta pelo autor organizados de acordo com os ataques. Fonte: Elaborado pelo autor.

Classificador	*Bruteforce*	*Infiltration*	DDoS	*Portscan*	*Botnet*	Web
Stackings	**0.017%**	**0.008%**	**0.053%**	**0.026%**	**0.196%**	**0.031%**
k-NN	0.018%	0.009%	0.085%	0.044%	0.282%	0.032%
DT	0.029%	0.011%	**0.053%**	0.031%	0.210%	0.049%
SVM	0.528%	0.131%	1.250%	0.464%	0.478%	1.256%
MLP	0.235%	0.131%	0.320%	0.408%	0.448%	0.161%

Attack (dataset)	**Exaustion (seconds)**	**Exaustion (hours)**
Brute-force	265145	73.6
Infiltration	62527	17.3
DDoS	361117	100.3
Portscan	415095	115.3
Botnet	67429	18.3
Web	104576	29.04

ALGUNS RESULTADOS INTERESSANTES...

IEEE SMC 2022
PRAGUE • CZECH REPUBLIC

ERROR RATE COMPARISON BETWEEN INDIVIDUAL CLASSIFIERS AND OUR APPROACH

Approach	Brute-force	Infiltration	DDoS	Portscan	Botnet	Web
Our	**0.018%**	**0.008%**	0.629%	**0.029%**	**0.202%**	**0.032%**
k-NN	**0.018%**	0.009%	0.085%	0.044%	0.282%	**0.032%**
DT	0.029%	0.011%	**0.053%**	0.031%	0.210%	0.049%
SVM	0.528%	0.131%	1.250%	0.464%	0.478%	1.256%
MLP	0.235%	0.131%	0.320%	0.408%	0.448%	0.161%

TABLE II
TIME RATIO (COMPUTATIONAL COST IN SECONDS AND HOURS) TO OBTAIN THE BEST COMMITTEES THROUGH EXHAUSTION, THROUGH DIVERSITY PRUNING AND THE DIFFERENCE (GAIN) BY OUR APPROACH.

Attack (dataset)	Exaustion (seconds)	Exaustion (hours)	Diversity (seconds)	Diversity (hours)	Difference (hours)
Brute-force	265145	73.6	8889	2.4	**–71.18**
Infiltration	62527	17.3	291	0.08	**–17,28**
DDoS	361117	100.3	15761	4.3	**–95.93**
Portscan	415095	115.3	19912	5.5	**–109.77**
Botnet	67429	18.3	684	0.19	**–18.54**
Web	104576	29.04	2272	0.6	**–28.41**

ALGUNS RESULTADOS INTERESSANTES...

IEEE SMC 2023
Maui, Hawaii

Approach	TN	FP	FN	TP	Sum of errors ↑
KDD-Cup'99					
Diversity	**97114**	**162**	**1473**	**395270**	**1635**
KNN	97047	229	1783	394960	2012
DT	96772	504	3032	393711	3536
MLP	96591	685	3800	392943	4485
SVM	96346	930	5222	391521	6152
NSL-KDD					
Diversity	**935**	**61**	**44**	**147476**	**105**
DT	910	86	57	147463	143
KNN	806	190	109	147411	299
MLP	759	237	133	147387	370
SVM	258	738	414	147106	1152
UNSW-NB15					
Diversity	63284	29716	4788	159885	**34504**
k-NN	50089	42911	**1853**	**162820**	44764
DT	**65134**	**27866**	19723	144950	47589
MLP	54096	38904	9741	154932	48645
SVM	62801	30199	31221	133452	61420
ISCX-IDS-2012					
Diversity Pruning	**20585**	**3550**	**491**	**756857**	**4041**
MLP	19008	5127	3351	753997	8478
KNN	20228	3907	11154	746194	15061
SVM	15452	8683	14458	742890	23141
DT	20346	3789	23080	734268	26869

ALGUNS RESULTADOS INTERESSANTES...

IEEE SMC 2023
Maui, Hawaii

Approach	Train time	Test time	Train gain	Test gain
		KDD-Cup'99		
Exaustion	267,435s	3m21s	–	–
Diversity	3,508s	231 *ms*	−98.69%	−99.89%
		NSL-KDD		
Exaustion	245,908s	2m33s	–	–
Diversity	2,542s	157 *ms*	−98.97%	−99.91%
		UNSW-NB15		
Exaustion	377,119s	6m07s	–	–
Diversity	6,004s	499 *ms*	−98.41%	−99.87%
		ISCX-IDS-2012		
Exaustion	658,546s	7m36s	–	–
Diversity	9,774s	576 *ms*	−98.41%	−99.86%

ALGUNS RESULTADOS INTERESSANTES…

31st IWSSIP – International Conference on Systems, Signals and Image Processing 2024

9th – 11th of JULY, 2024, IEEE Technical Co-Sponsored Conference

Graz University of Technology, Inffeldgasse 12, A-8010 Graz, AUSTRIA,

Approach	Accuracy ↓	AUC	Precision	Recall	F1-Score
KDD-Cup'99					
Diversity	**99.669%**	**99.731%**	**99.959%**	**99.629%**	**99.794%**
k-NN	99.593%	99.658%	99.942%	99.551%	99.746%
DT	99.284%	99.359%	99.872%	99.236%	99.553%
MLP	99.092%	99.169%	99.826%	99.042%	99.433%
SVM	98.755%	98.864%	99.763%	98.684%	99.220%
NSL-KDD					
Diversity	**99.929%**	**96.923%**	**99.959%**	**99.970%**	**99.964%**
DT	99.904%	95.663%	99.942%	99.961%	99.952%
k-NN	99.799%	90.425%	99.871%	99.926%	99.899%
MLP	99.751%	88.057%	99.839%	99.910%	99.875%
SVM	99.224%	62.811%	99.501%	99.719%	99.610%
UNSW-NB15					
Diversity	**86.609%**	**82.570%**	**84.327%**	97.092%	**90.261%**
k-NN	82.628%	76.367%	79.142%	**98.875%**	87.915%
DT	81.531%	79.030%	83.875%	88.023%	85.899%
MLP	81.121%	76.126%	79.929%	94.085%	86.431%
SVM	76.164%	74.284%	81.547%	81.041%	81.293%
ISCX-IDS-2012					
Diversity	**99.483%**	**92.613%**	**99.533%**	**99.935%**	**99.734%**
MLP	98.915%	89.157%	99.325%	99.558%	99.441%
k-NN	98.073%	91.170%	99.479%	98.527%	99.001%
SVM	97.039%	81.057%	98.845%	98.091%	98.466%
DT	96.562%	90.627%	99.487%	96.953%	98.203%

TU Graz

ALGUNS RESULTADOS INTERESSANTES…

LATINOWARE 2025

22º Congresso Latino-americano de Software Livre e Tecnologias Abertas

Latin.Science 2025

Machine Learning-based Spyware Detection Systems: An Undersampling Performance Analysis

Large Language Models for Intrusion Detection: Tokenization Impacts on DDoS Flows

Interpretability of Intrusion Detection Models: An Information Visualization Approach

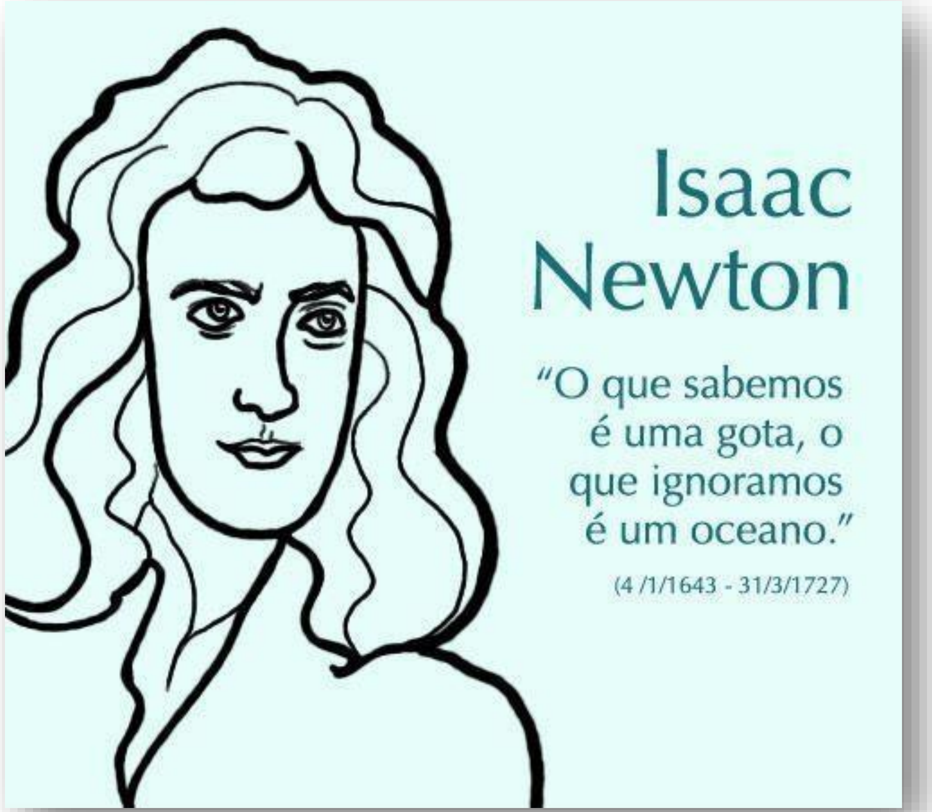
Evaluation of Machine Learning Algorithms for Intrusion Detection in SCADA Systems

Detection of SQL Injection: A Comparative Analysis of Machine Learning and Deep Learning Algorithms

Comparative Evaluation of Supervised Machine Learning Algorithms for Intrusion Detection in Electric Vehicle Systems

QUAIS AS CONCLUSÕES?

- Os ataques evoluíram muito
    - Pentest é muuuuito mais atrativo
- Blue Team precisa evoluir
    - I.A. aplicada a segurança defensiva é realidade!
- Red Team pode pensar nas IAs generativas ☺



OBRIGADO!

Thiago José Lucas
Doutor em Ciência da Computação | Especialista em Redes de Computadores e Cibersegurança Defensiva
Ourinhos, São Paulo, Brazil · Contact info