Inteligência Artificial e Cibersegurança

Outubro—Mês Europeu da Cibersegurança

Diogo Nuno Freitas
diogo.freitas@iti.larsys.pt

Ciclo de Conversas *online* com quem sabe.

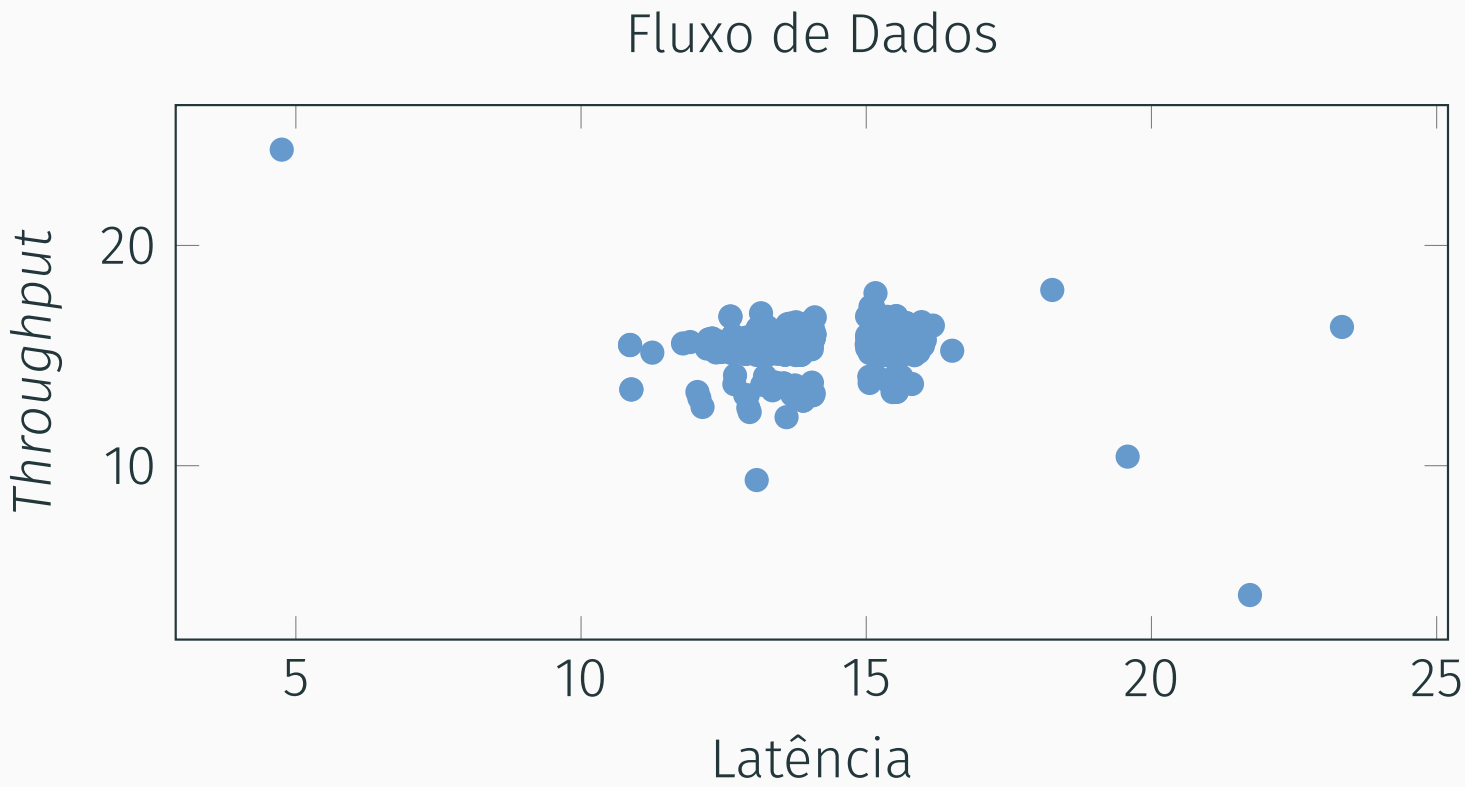
1 de novembro de 2021

EUROPE DIRECT
Madeira

Motivação

Motivação

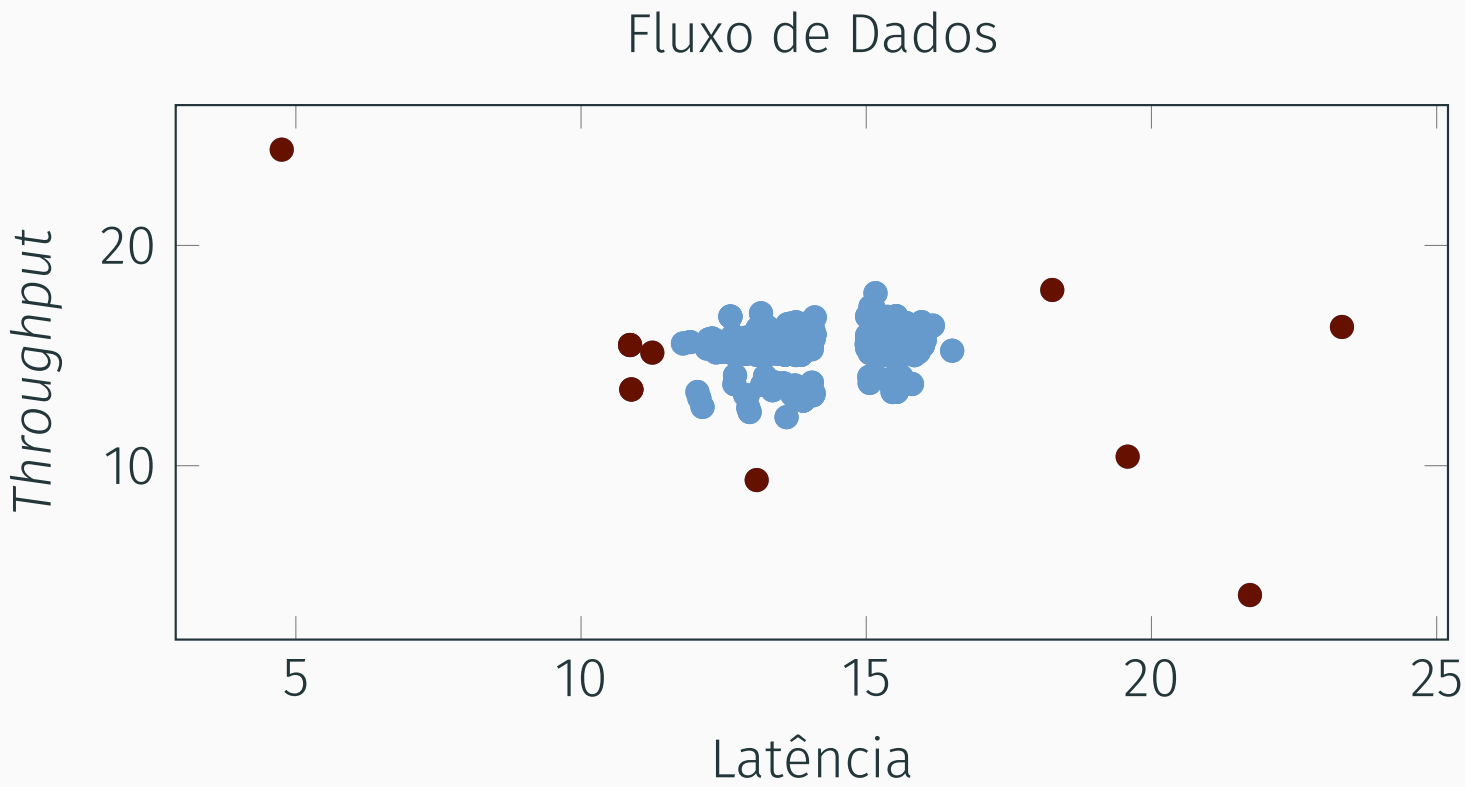
Olhando para este gráfico, quais os pontos que nos **chamam mais a atenção?**



Como pode um sistema detectar automaticamente esses pontos?

Motivação

Olhando para este gráfico, quais os pontos que nos **chamam mais a atenção?**



Como pode um sistema detectar automaticamente esses pontos?

Motivação (cont.)

Podemos começar por calcular a **média** e o **desvio padrão**, isto é,

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \approx 14,42 \qquad \text{e} \qquad \sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \approx 1,45.$$

Vamos usar essas duas medidas para calcular o **limite** (l) para os dados considerados *normais*:

$$l = \sigma \times c,$$

onde c é uma constante arbitrária.

Motivação (cont.)

Finalmente, podemos definir os limites **inferiores e superiores**, de tal forma que:

$$\begin{cases} l_{\text{inf}} = \bar{x} - l, \\ l_{\text{sup}} = \bar{x} + l. \end{cases}$$

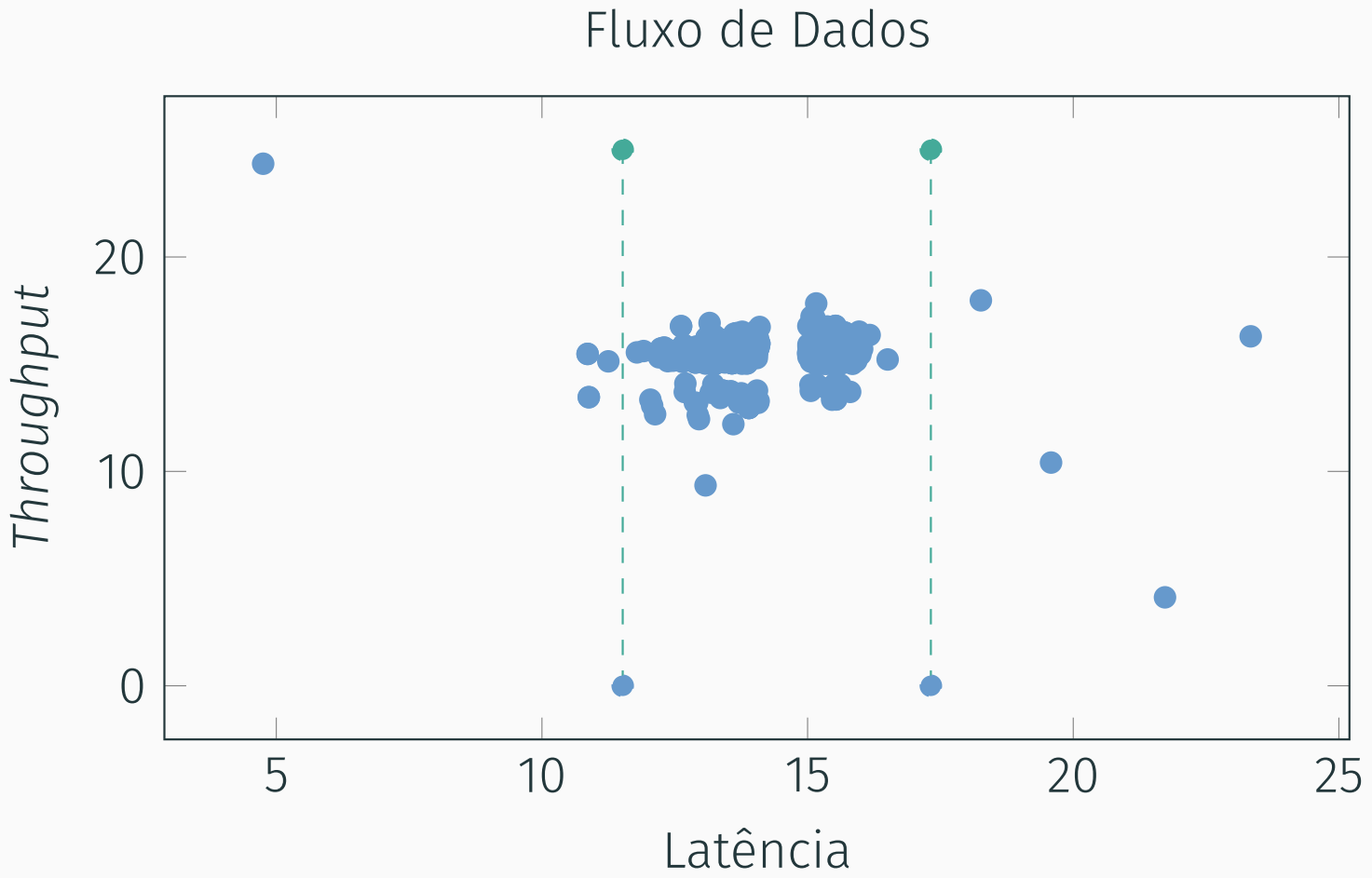
Ou seja, um **ponto é considerado *normal* se** $l_{\text{inf}} \leq x_i \leq l_{\text{sup}}$.

Para o nosso caso, e considerando $c = 2$,

$$\begin{cases} l_{\text{inf}} = 14,42 - 2,90 \approx 11,52, \\ l_{\text{sup}} = 14,42 + 2,90 \approx 17,32. \end{cases}$$

Motivação (cont.)

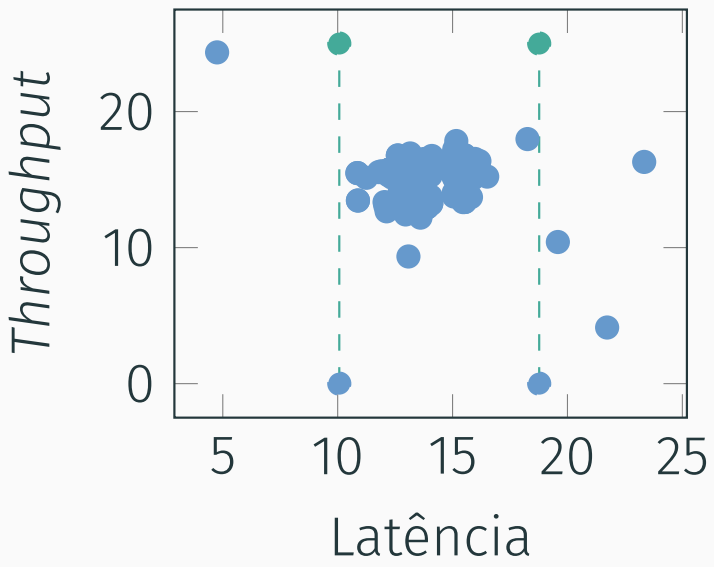
Graficamente, estamos a **criar duas rectas verticais**, tais como,



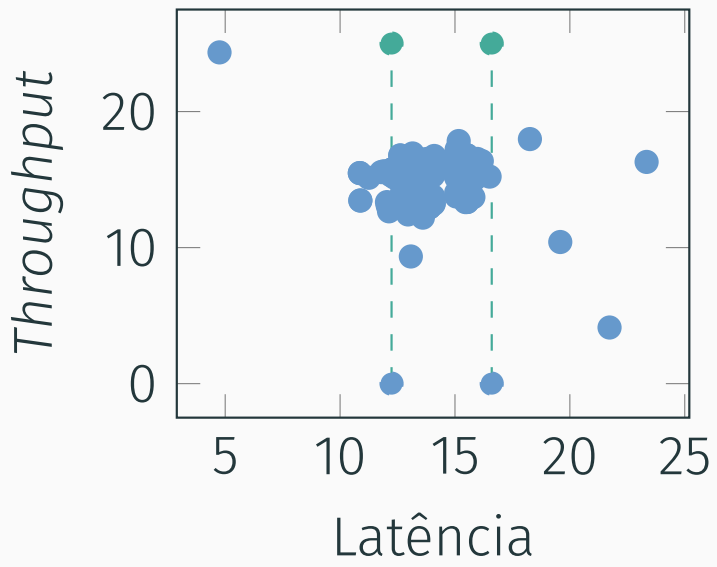
Motivação (cont.)

Como definir c?

Fluxo de Dados ($c = 3$)



Fluxo de Dados ($c = 1,5$)



Motivação: Como definir o valor de c?

Precisamos de uma forma automática (e adaptativa) para classificar os pontos como sendo *normais* ou *outliers* $\rightarrow$ **Inteligência Artificial**.

Objectivos da sessão

Objectivos da sessão

1. Perceber a motivação para o uso de **métodos automáticos de classificação**.
2. Ter uma percepção breve sobre a **história da Inteligência Artificial** (IA).
3. Introduzir **conceitos básicos** de *machine learning* (ML).
4. Conhecer os **modelos de IA utilizados actualmente**.
5. Perceber como é que esses modelos podem ser **aplicados no âmbito da cibersegurança**.

Introdução

O que é a IA?

É uma **invenção humana** que permite usar sistemas informáticos para **simular comportamentos humanos inteligentes**.

Com a IA, assistimos **a uma mudança no paradigma de programação** que até então estávamos habituados.

Programação tradicional:

Machine Learning:

Introdução (cont.)

Acontecimentos importantes:

- Em 1943, **McCulloch e Pitts** desenvolveram o **primeiro modelo computacional inteligente**: uma rede neuronal artificial.
- Rosenblatt nos finais da década de 50 e começo da década de 60, sugere **perceptrões**.
- Linnainmaa sugere o algoritmo de de *backpropagation* (1970).
- Em 1979, Fukushima propõe as **rede neuronais convolucionais** (CNN).

Introdução (cont.)

Outros acontecimentos importantes:

- As redes ***long short-term memory*** **(LSTM)** são sugeridas em 1997.
- Na década de 2000, começam a surgir implementações dos **algoritmos de treino em GPUs**.
- Goodfellow e colaboradores introduzem as redes ***Generative Adversarial*** em 2014.
- Recentemente, o conceito de ***transformers*** **foi introduzido** para processamento de linguagem natural.

IA e a cibersegurança

Já existem diversas aplicações de modelos de IA à cibersegurança. E existirão cada vez mais!

Vantagens:

- Possibilita **detectar padrões em milhões tuplos de dados**.
- Permite fazer **processamento de linguagem natural**.
- Possibilita executar aplicações que estão sempre ***atentas***.
- Garante que os modelos **aprendam constantemente com novas informações**.

IA e a cibersegurança (cont.)

Existem, contudo, algumas dificuldades que estão **dependentes dos humanos**.

Dificuldades:

- **Elevado desequilíbrio** entre o número de dados de ameaças e os de não-ameaça.
- O **sucesso dos modelos está sempre dependente da disponibilização de informação**.
- **Falta de recursos humanos** qualificados na área de IA.
- **Custos elevados** com formação e com equipamentos.
- Pode acontecer virar-se o **feitiço contra o feiticeiro**!

Conceitos básicos

Conceitos básicos

O que são *features*?

Features ou *atributos* são propriedades que **descrevem uma possível relação com a variável de saída**. Os atributos podem ser numéricos ou categóricos.

Criação de *features*

Métodos manuais ou automáticos de **criação de novos atributos com base nos que já existem**.

Selecção de *features*

Métodos manuais ou automáticos para **reduzir a dimensionalidade do problema**.

Conceitos básicos (cont.)

Quais são os dois tipos de problemas mais comuns para ML?

Classificação

Os modelos utilizam os atributos fornecidos para **categorizar os dados em classes**. Maioritariamente associada com tarefas de *clustering*.

Regressão

Os modelos tentam **inferir uma função real** para mapear os atributos e as variáveis de saída (em valor real).

Conceitos básicos (cont.)

Quais são os dois tipos de aprendizagem mais comuns para ML?

Supervised learning

Quando disponibilizamos ao modelo **exemplos de mapeamento entre as entradas (atributos) e as saídas**.

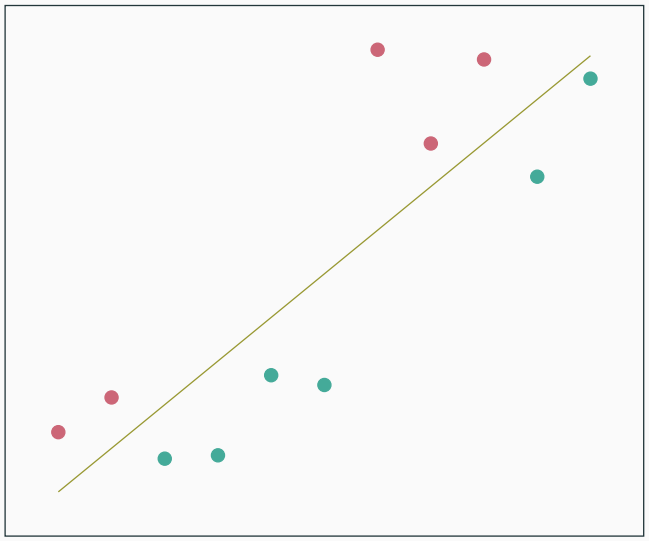
Unsupervised learning

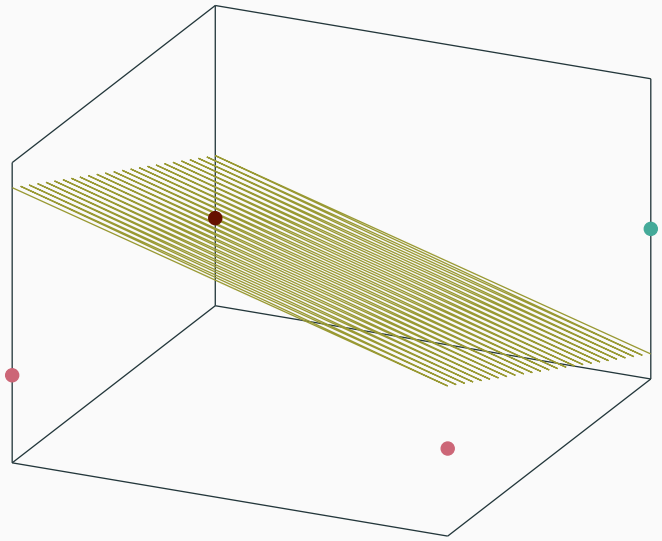
No caso do ***unsupervised learning***, esse mapeamento **não é fornecido à rede**. A rede é, assim, obrigada **a aprender a classificar os dados com base nos padrões ou nos *clusters*** que encontrar.

Exemplos práticos

Objectivo de modelo

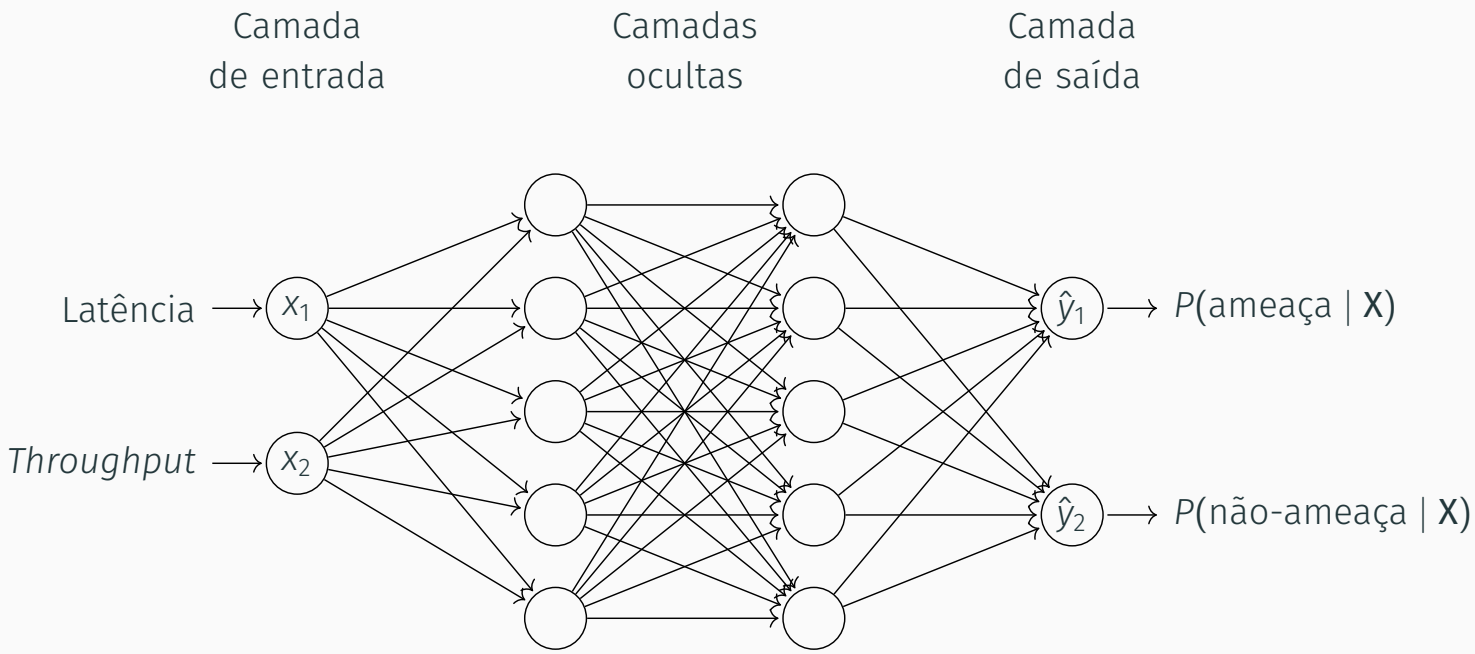
Na sua essência, o faz um modelo de ML?





Num problema de classificação, tenta encontrar **uma linha ou um plano que permita separar as classes**.

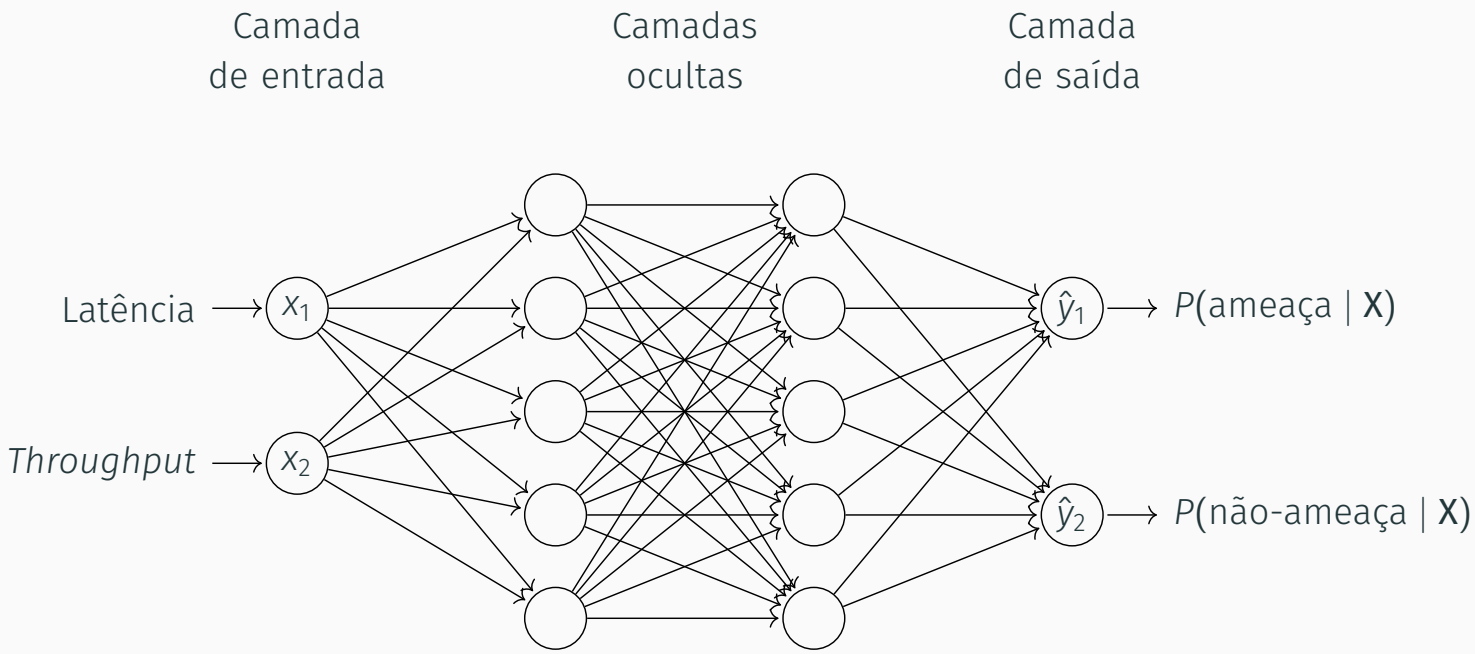
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

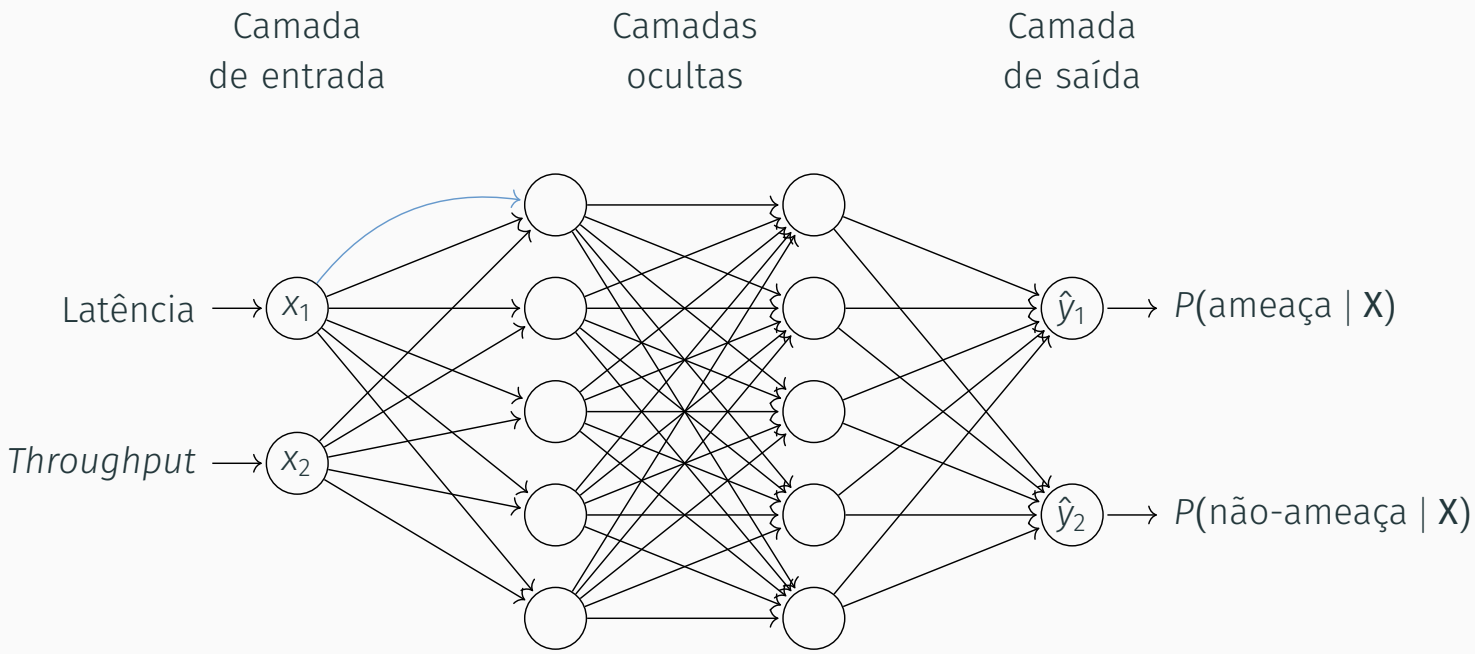
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

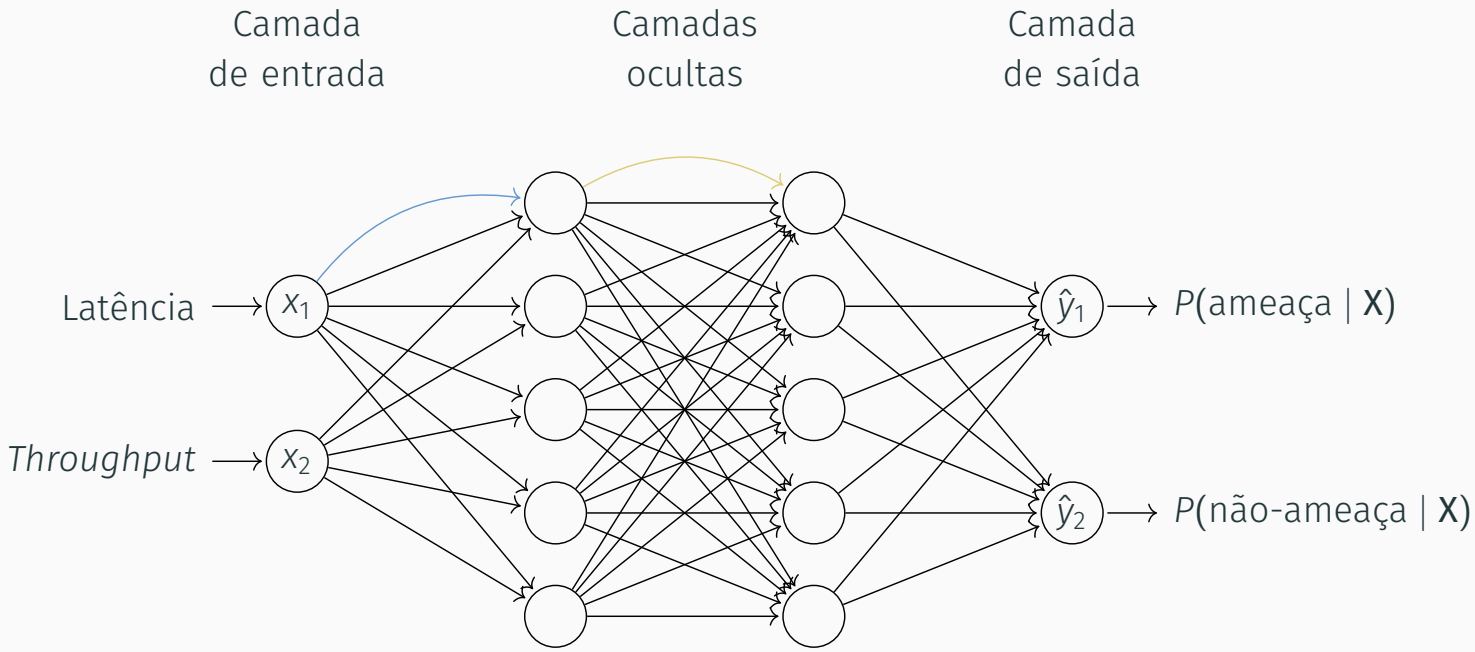
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

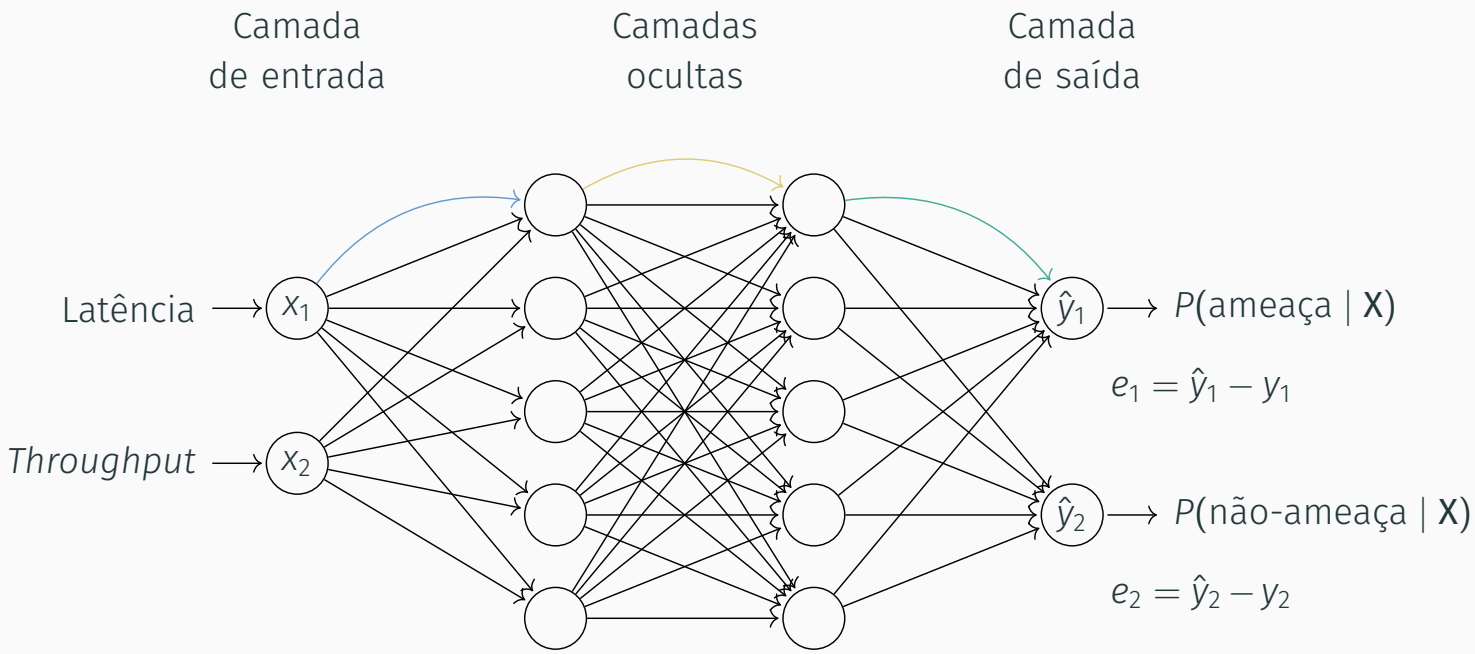
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

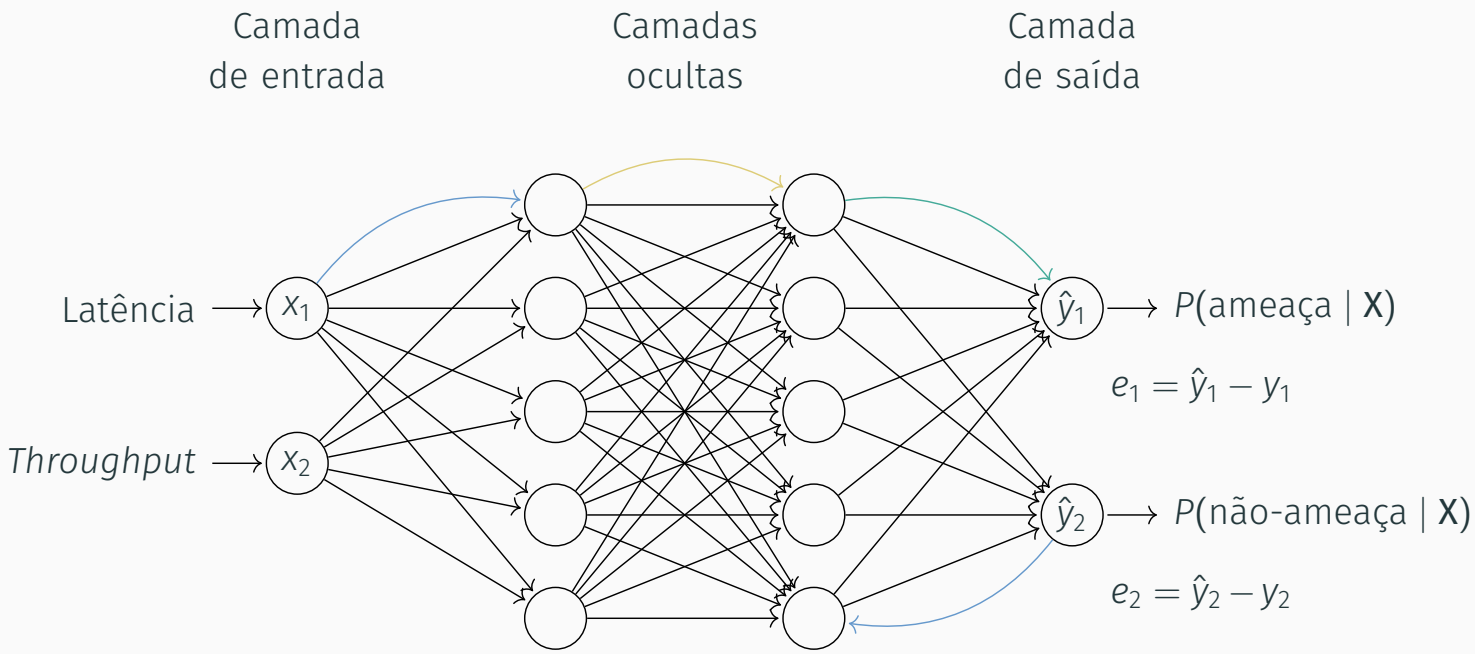
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

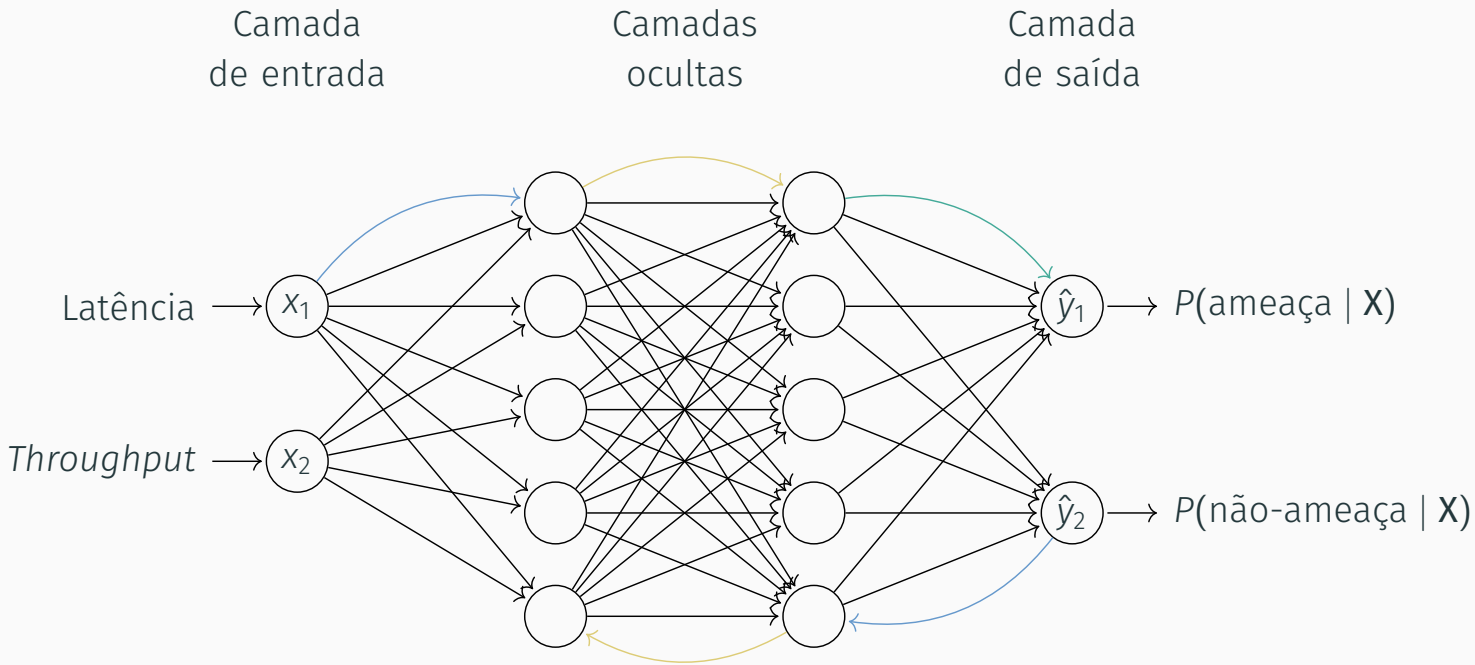
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

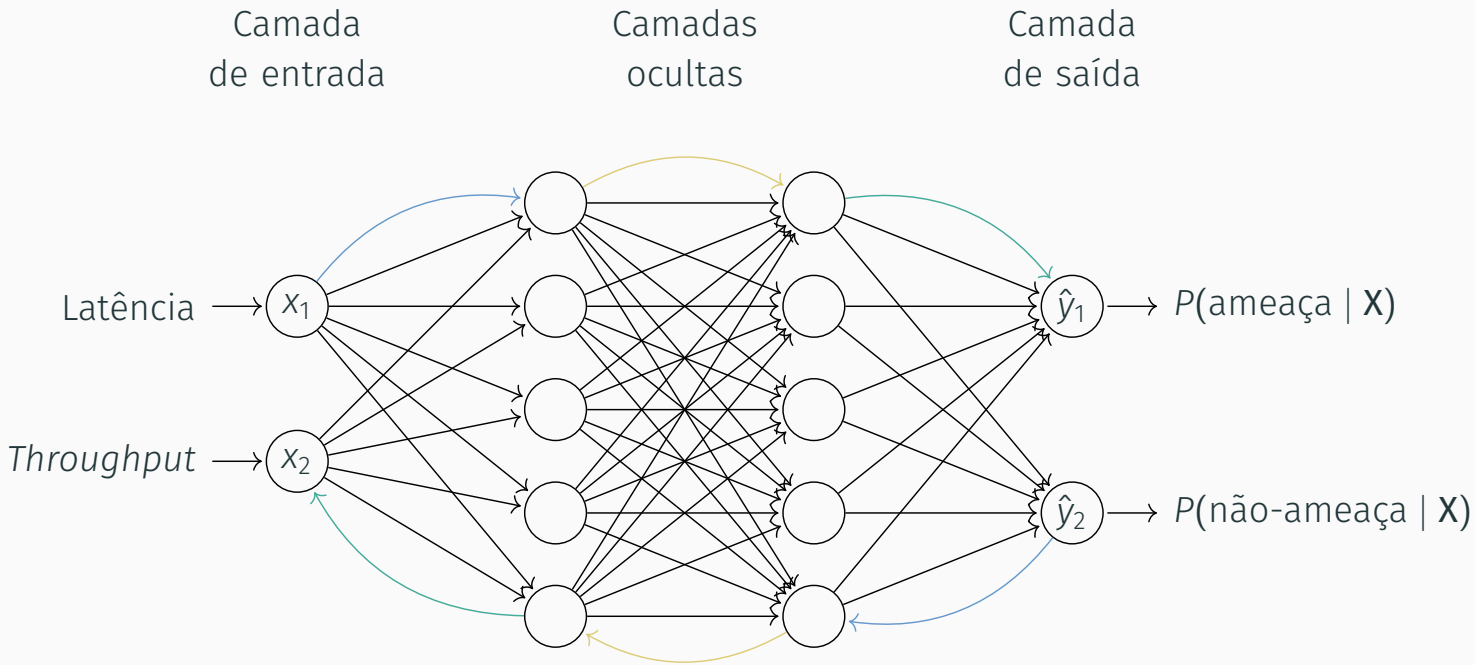
Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

Exemplo 1: Redes Neuronais Artificiais



Numa rede neuronal aritificial (RNA), existem duas passagens:

- *Forward Propagation*.
- *Backward Propagation*.

Exemplo 1: Redes Neuronais Artificiais (cont.)

Os neurónios numa **RNA têm duas funções**:

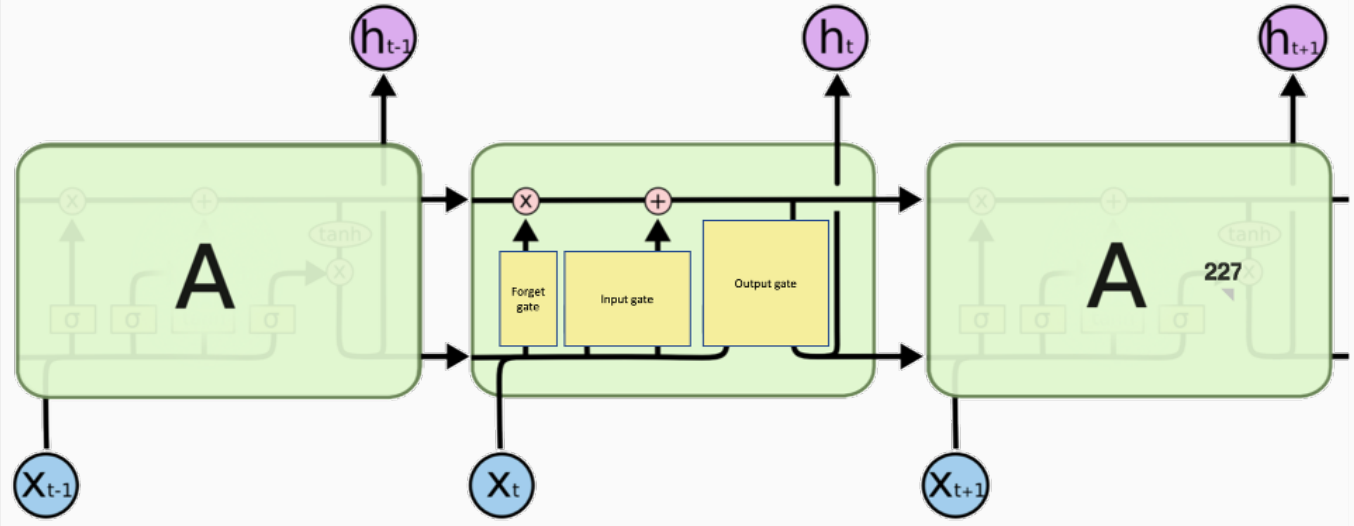
- **Função de transferência** ($\mathsf{T} = \mathsf{W} \cdot \mathsf{I}$);
- **Função de activação** (ϕ).

O **algoritmo de *backpropagation* consiste em dois aspectos** fundamentais:

- **Cálculo do erro de saída e retropropagação** até aos neurónios de entrada;
- **Actualização dos pesos** por meio de um algoritmo de optimização.

Exemplo 2: LSTMs

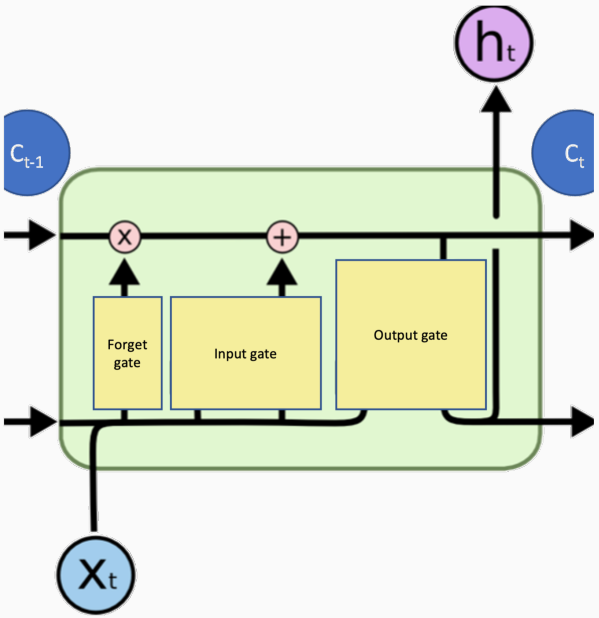
A **estrutura de uma rede Long Short Term Memory (LSTM)** é a seguinte[1]:



As LSTMs são modelos úteis quando existe uma **sequência ordenada nos dados**, pois conseguem guardar **contexto**.

[1] Imagem adaptada de: https://colah.github.io/posts/2015-08-Understanding-LSTMs

Exemplo 2: LSTMs (cont.)



Cada célula de uma **LSTM possui três operações/portas** (*gates*):

- ***forget***—que informação esquecer?
- ***input***—o que adicionar ao que já sei?
- ***output***—que informação passar?

Exemplo 2: LSTMs (cont.)

Este tipo de redes consegue aprender **dependências douradoras, bem como esporádicas.**

No âmbito da cibersegurança, as LSTMs podem ser usadas **para detectar e prevenir fraudes**. Vejamos um exemplo.

t	x	y
1	15	0
2	45	0
3	65	0
4	500	1
5	500	1
6	25	0

Exemplo 2: LSTMs (cont.)

Para dar contexto à LSTM, os dados **precisam de reestruturados em janela deslizante**. Neste caso, o tamanho da janela escolhido foi 3.

x_1	x_2	x_3	y
15	45	65	0
45	65	500	0
65	500	500	1
500	500	25	1

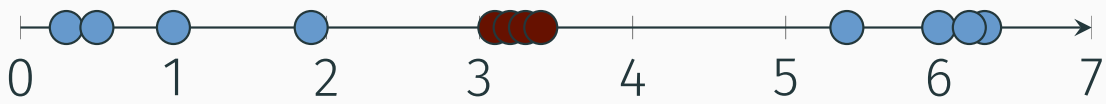
Iremos ter uma LSTM **com três células para predizer** y, e detectar movimentos potencialmente fraudulentos.

Exemplo 3: SVMs e NLP

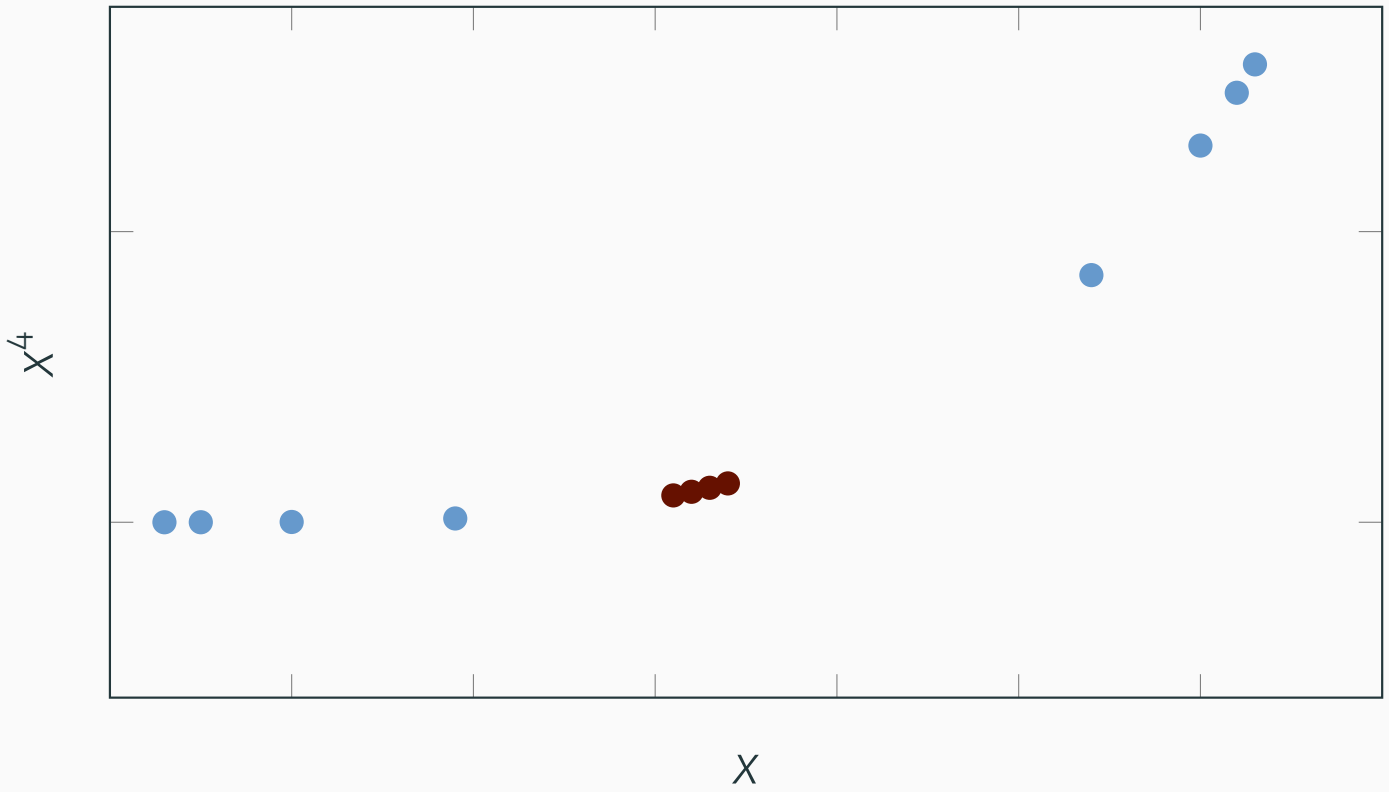
O último exemplo que vamos ver usa as **Support Vector Machines (SVMs)** e o **Processamento de Linguagem Natural** (NLP).

Comecemos por entender como é que as SVMs funcionam.

Como podemos **separar as seguintes classes**?

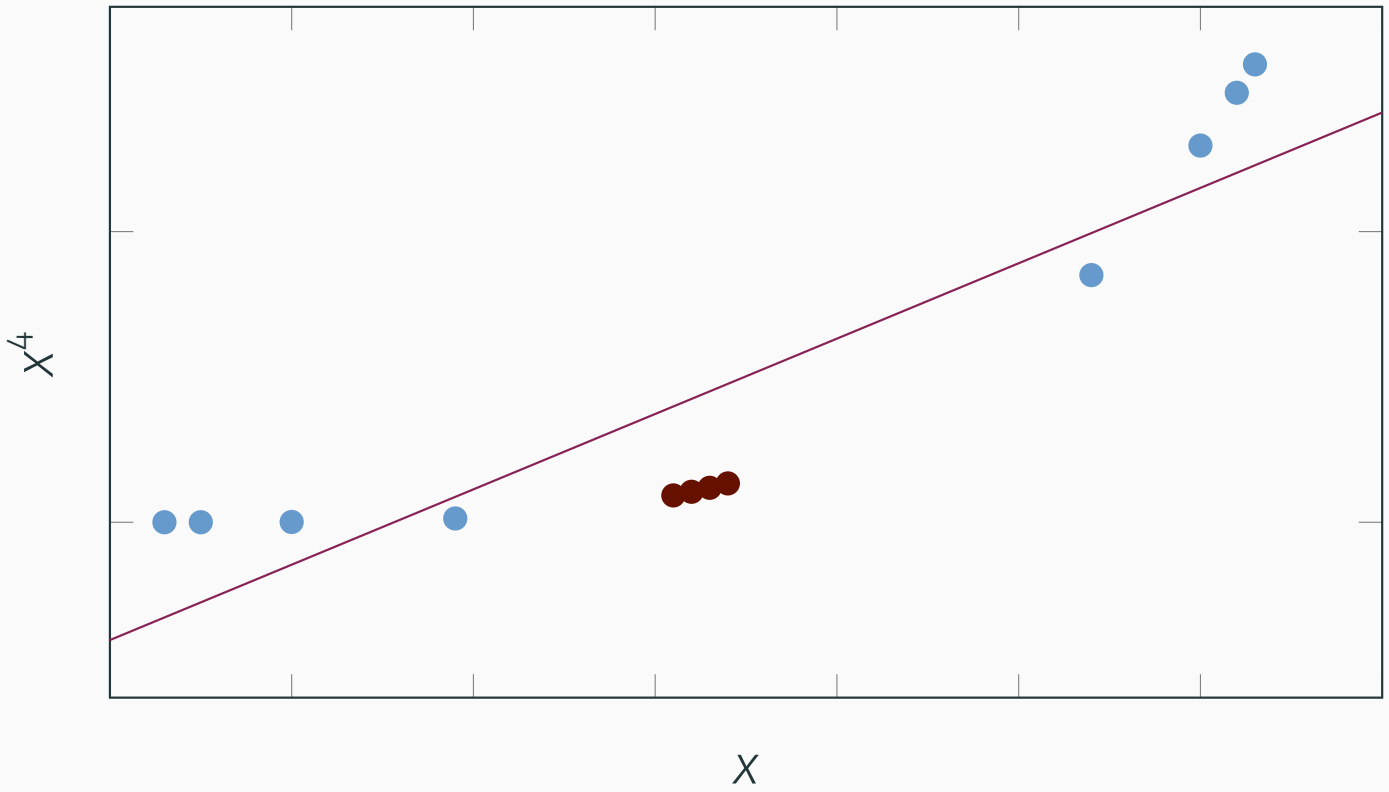


Exemplo 3: SVMs e NLP (cont.)



As SVMs na sua essência encontram um **hiperplano capaz de separar as duas classes**.

Exemplo 3: SVMs e NLP (cont.)



As SVMs na sua essência encontram um **hiperplano capaz de separar as duas classes**.

Exemplo 3: SVMs e NLP (cont.)

Como podemos desenvolver um sistema automático para detectar se um e-mail é *SPAM*, ou não?

Abordagem *n*-grama

Permite **separar o texto em sequências** de *n* caracteres ou palavras.

Modelo *bag-of-words*

Permite **transformar a linguagem natural em representação numérica**. A representação numérica é feita à custa da **frequência de cada termo**.

A partir daqui, **passamos a ter um problema comum de ML** que podemos resolver usando as SVMs.

Conclusão

Conclusão

Começámos por ver **como podemos classificar manualmente** acontecimentos anormais.

Percebemos **como surgiram os primeiros modelos de ML**.

Entendemos o que são **atributos**, **quais os problemas mais comuns** e que **tipos de aprendizagem existem**.

Finalmente, **aplicámos três modelos de ML** a problemas de cibersegurança.

Conclusão (cont.)

	ANN
Tipos de problema	Classificação e regressão.
Tipos de aprendizagem	*Supervised*.
Dados	Qualquer tipo de dados.
Exemplos de aplicação	Dados não ordenados (geral).

	LSTM
Tipos de problema	Classificação e regressão.
Tipos de aprendizagem	*Supervised*.
Dados	Bases de dados com muitos registos.
Exemplos de aplicação	Previsão (séries ordenadas).

Conclusão (cont.)

	SVM
Tipos de problema	Classificação e regressão.
Tipos de aprendizagem	*Supervised*.
Dados	Dados de grandes dimensões.
Exemplos de aplicação	NLP.

Bibliografia i

[1] S. Aiyar and N. P. Shetty.
N-gram assisted youtube spam comment detection.
Procedia Computer Science, 132:174–182, 2018.
International Conference on Computational Intelligence and Data Science.

[2] A. Parisi.
Hands-On Artificial Intelligence for Cybersecurity: Implement Amart AI Systems for Preventing Cyber Attacks and detecting threats and network anomalies.
Packt Publishing, 2019.

[3] C. Sammut and G. Webb.
Encyclopedia of Machine Learning.
Encyclopedia of Machine Learning. Springer US, 2011.

Bibliografia ii

[4] E. Tsukerman.
Machine Learning for Cybersecurity Cookbook: Over 80 Recipes on How to Implement Machine Learning Algorithms for Building Security Systems Using Python.
Packt Publishing, 2019.

Inteligência Artificial e Cibersegurança

Outubro—Mês Europeu da Cibersegurança

Diogo Nuno Freitas
diogo.freitas@iti.larsys.pt

Ciclo de Conversas *online* com quem sabe.

1 de novembro de 2021

EUROPE DIRECT
Madeira