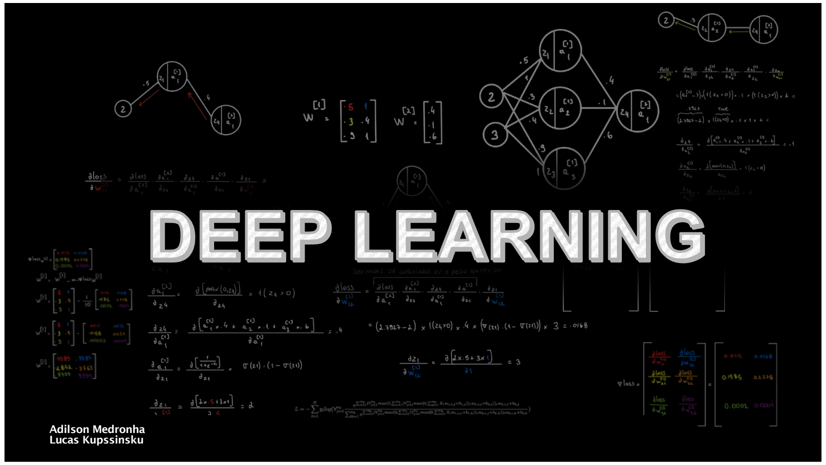
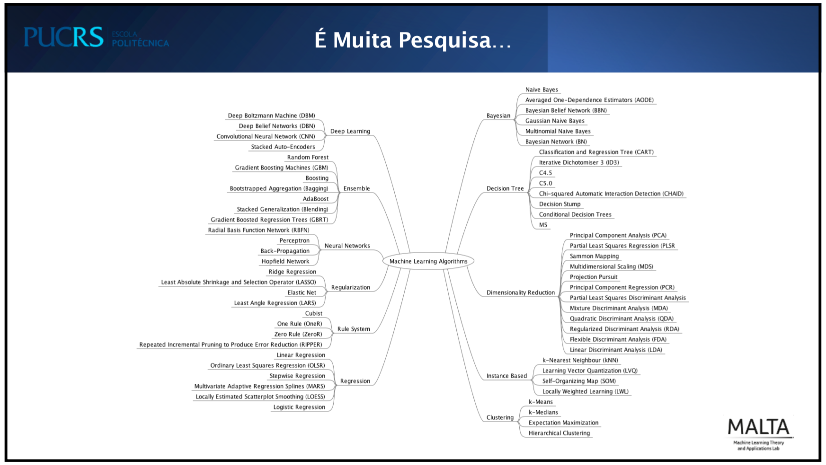


DEEP LEARNING
Adilson Medronha
Lucas Kupssinsku

1

APRENDIZADO DE MÁQUINA
O QUE É LEARNING?
PRIMEIRO ALG DE LEARNING
PROTOCOLOS DE TREINAM.
OTIMIZAÇÃO
REDES NEURAIS
BACKPROP + PRATICA
CNN + PRATICA

2



PUCRS ESCOLA POLITÉCNICA
É Muita Pesquisa…
Machine Learning Algorithms
Deep Learning
Deep Boltzmann Machine (DBM)
Deep Belief Networks (DBN)
Convolutional Neural Network (CNN)
Stacked Auto-Encoders
Ensemble
Random Forest
Gradient Boosting Machines (GBM)
Boosting
Bootstrapped Aggregation (Bagging)
AdaBoost
Stacked Generalization (Blending)
Gradient Boosted Regression Trees (GBRT)
Neural Networks
Radial Basis Function Network (RBFN)
Perceptron
Back-Propagation
Hopfield Network
Regularization
Ridge Regression
Least Absolute Shrinkage and Selection Operator (LASSO)
Elastic Net
Least Angle Regression (LARS)
Rule System
Cubist
One Rule (OneR)
Zero Rule (ZeroR)
Repeated Incremental Pruning to Produce Error Reduction (RIPPER)
Regression
Linear Regression
Ordinary Least Squares Regression (OLSR)
Stepwise Regression
Multivariate Adaptive Regression Splines (MARS)
Locally Estimated Scatterplot Smoothing (LOESS)
Logistic Regression
Bayesian
Naive Bayes
Averaged One-Dependence Estimators (AODE)
Bayesian Belief Network (BBN)
Gaussian Naive Bayes
Multinomial Naive Bayes
Bayesian Network (BN)
Decision Tree
Classification and Regression Tree (CART)
Iterative Dichotomiser 3 (ID3)
C4.5
C5.0
Chi-squared Automatic Interaction Detection (CHAID)
Decision Stump
Conditional Decision Trees
M5
Dimensionality Reduction
Principal Component Analysis (PCA)
Partial Least Squares Regression (PLSR
Sammon Mapping
Multidimensional Scaling (MDS)
Projection Pursuit
Principal Component Regression (PCR)
Partial Least Squares Discriminant Analysis
Mixture Discriminant Analysis (MDA)
Quadratic Discriminant Analysis (QDA)
Regularized Discriminant Analysis (RDA)
Flexible Discriminant Analysis (FDA)
Linear Discriminant Analysis (LDA)
Instance Based
k-Nearest Neighbour (kNN)
Learning Vector Quantization (LVQ)
Self-Organizing Map (SOM)
Locally Weighted Learning (LWL)
Clustering
k-Means
k-Medians
Expectation Maximization
Hierarchical Clustering
MALTA
Machine Learning Theory and Applications Lab

3

4

PUCRS ESCOLA POLITÉCNICA

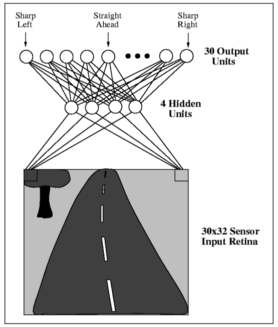
MOTIVAÇÃO

ALVINN:
AN AUTONOMOUS LAND VEHICLE IN A NEURAL NETWORK

Dean A. Pomerleau
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

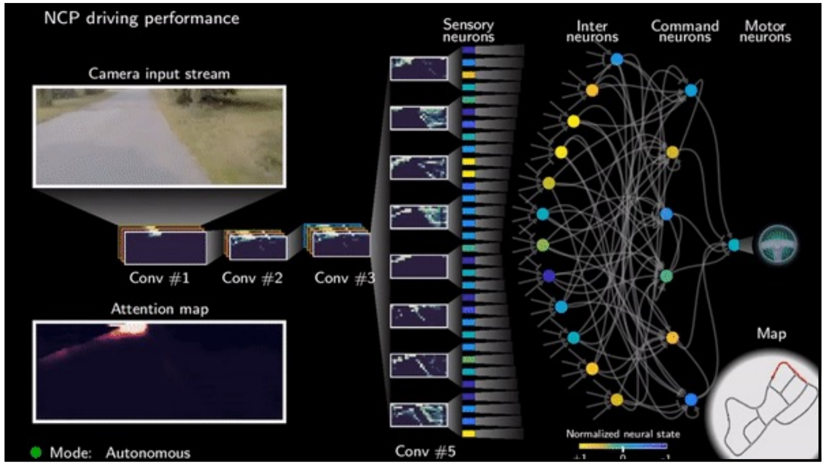
ABSTRACT

ALVINN (Autonomous Land Vehicle In a Neural Network) is a 3-layer back-propagation network designed for the task of road following. Currently ALVINN takes images from a camera and a laser range finder as input and produces as output the direction the vehicle should travel in order to follow the road. Training has been conducted using simulated road images. Successful tests on the Carnegie Mellon autonomous navigation test vehicle indicate that the network can effectively follow real roads under certain field conditions. The representation developed to perform the task differs dramatically when the network is trained under various conditions, suggesting the possibility of a novel adaptive autonomous navigation system capable of tailoring its processing to the conditions at hand.



MALTA

5



6

PUCRS ESCOLA POLITÉCNICA

MOTIVAÇÃO

MALTA

7

PUCRS ESCOLA POLITÉCNICA

MOTIVAÇÃO

A Method for Animating Children's Drawings of the Human Figure

HARRISON JESSE SMITH, Meta AI Research, USA
QINGYUAN ZHENG*, Tencent America, USA
YIFEI LI*, MIT CSAIL, USA
SOMYA JAIN, Meta AI Research, USA
JESSICA K. HODGINS*, Carnegie Mellon University, USA

MALTA

8

PUCRS ESCOLA POLITÉCNICA

MOTIVAÇÃO

A Method for Animating Children's Drawings of the Human Figure

HARRISON JESSE SMITH, Meta AI Research, USA
QINGYUAN ZHENG*, Tencent America, USA
YIFEI LI*, MIT CSAIL, USA
SOMYA JAIN, Meta AI Research, USA
JESSICA K. HODGINS*, Carnegie Mellon University, USA

Referência

MALTA Machine Learning Theory and Applications Lab

9

PUCRS ESCOLA POLITÉCNICA

MOTIVAÇÃO

MALTA Machine Learning Theory and Applications Lab

10

PUCRS ESCOLA POLITÉCNICA

MOTIVAÇÃO

"a storm trooper vacuuming a beach"

"sunset time lapse at the beach with moving clouds and colors in the sky, 4k, high resolution"

"an astronaut feeding ducks on a sunny afternoon, reflection from the water"

MALTA Machine Learning Theory and Applications Lab

11

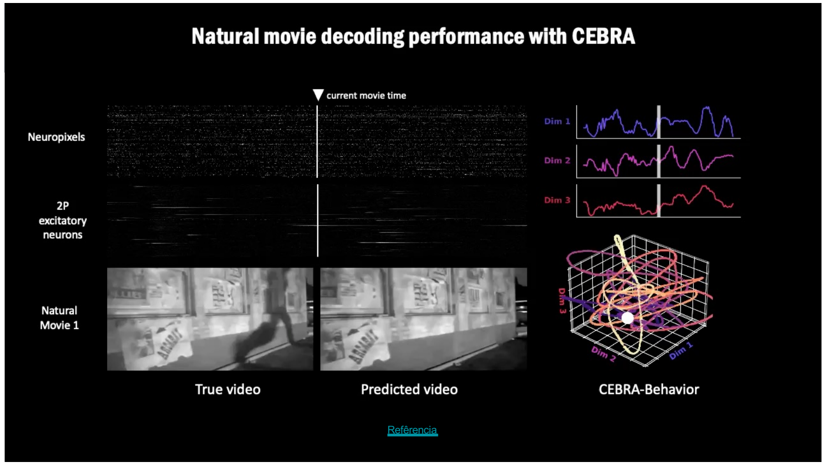
Referência

12

13


PUCRS
MOTIVAÇÃO
Refência
MALTA

14


Natural movie decoding performance with CEBRA
current movie time
Neuropixels
2P excitatory neurons
Natural Movie 1
Dim 1
Dim 2
Dim 3
True video
Predicted video
CEBRA-Behavior
Referência

15


APRENDIZADO DE MÁQUINA
> O QUE É LEARNING?
PRIMEIRO ALG DE LEARNING
PROTOCOLOS DE TREINAM.
OTIMIZAÇÃO
REDES NEURAIS
BACKPROP + PRATICA
CNN + PRATICA

16

“Na medida que os computadores se tornam mais sofisticados, parece inevitável que o Aprendizado de Máquina exerça um papel central em Ciência da Computação e tecnologia de computadores”
- Tom Mitchell

17

PUCRS ESCOLA POLITÉCNICA

O Que é Learning?

*“A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E.”* - Mitchell, 1997

MALTA Machine Learning Theory and Applications Lab

18

PUCRS ESCOLA POLITÉCNICA

O Que é Learning?

*“A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E.”* - Mitchell, 1997

- Experiência (E): **conjunto de dígitos anotados à mão (Hello World!)**

MALTA Machine Learning Theory and Applications Lab

19

PUCRS ESCOLA POLITÉCNICA

O Que é Learning?

*“A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E.”* - Mitchell, 1997

- Experiência (E): **conjunto de dígitos anotados à mão (Hello World!)**

- Tarefa (T): **classificação**

MALTA Machine Learning Theory and Applications Lab

20

O Que é Learning?

*"A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E."* - Mitchell, 1997

- Experiência (E): **conjunto de dígitos anotados à mão (Hello World!)**

- Tarefa (T): **classificação**

O que vocês acham, é viável programar uma solução imperativa?

- Medida de desempenho (P): **acurácia**

MALTA

21

O Que é Learning?

*"A computer program is said to **learn** from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in T, as measured by P, improves with experience E."* - Mitchell, 1997

- Experiência (E): **conjunto de dígitos anotados à mão (Hello World!)**

- Tarefa (T): **classificação**

O que vocês acham, é viável programar uma solução imperativa?

Não! Mas com redes neurais é **trivial**

- Medida de desempenho (P): **acurácia**

MALTA

22

O Que é Learning?

Learning = Representation + Evaluation + Optimization

MALTA

23

O Que é Learning?

Learning = Representation + Evaluation + Optimization

Representation

- Instances
  - K-nearest neighbor
  - Support vector machines
- Hyperplanes
  - Naive Bayes
  - Logistic regression
- Decision trees
- Sets of rules
  - Propositional rules
  - Logic programs
- NEURAL NETWORKS
- Linear programming
- Graphical models
  - Bayesian networks
  - conditional random fields

Evaluation

- ACCURACY/error rate
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Information gain
- K-l divergence
- Cost/Utility
- Margin

Optimization

- Combinatorial optimization
  - Greedy search
  - Beam search
  - Branch-and-bound
- Continuous optimization
  - Unconstrained
    - GRADIENT DESCENT
    - Conjugate gradient
    - Quasi-newton methods

MALTA

24

PUCRS ESCOLA POLITÉCNICA

Learning, Overview

CONJUNTO DE DADOS

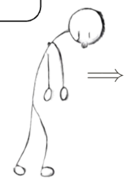
MALTA Machine Learning Theory and Applications Lab

25

PUCRS ESCOLA POLITÉCNICA

Learning, Overview

CONJUNTO DE DADOS



$\Longrightarrow$

x_1 pressão arterial

x_2 temperatura

x_3 peso

x_4 sente fadiga

x_5 tem tosse

$\vdots$

x_d etc

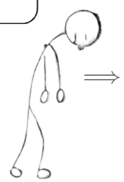
MALTA Machine Learning Theory and Applications Lab

26

PUCRS ESCOLA POLITÉCNICA

Learning, Overview

CONJUNTO DE DADOS



$\Longrightarrow$

x_1 pressão arterial

x_2 temperatura

x_3 peso

x_4 sente fadiga

x_5 tem tosse

$\vdots$

x_d etc

$\mathbf{x}^{(m)} = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$

$\mathbf{y}^{(m)} \in \{0, 1\}$

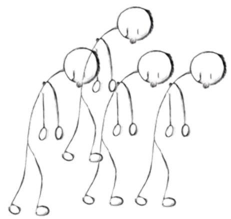
MALTA Machine Learning Theory and Applications Lab

27

PUCRS ESCOLA POLITÉCNICA

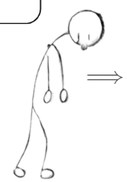
Learning, Overview

CONJUNTO DE DADOS



$$\underbrace{\begin{bmatrix} y^{(1)} & \cdots & y^{(m)} \end{bmatrix} \begin{bmatrix} x_1^{(1)} & \cdots & x_1^{(m)} \\ x_2 & \cdots & x_2 \\ \vdots & \vdots & \vdots \\ x_d & \cdots & x_d \end{bmatrix}}_{\text{Conjunto de dados}}$$

$\{((\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), (\mathbf{x}^{(2)}, \mathbf{y}^{(2)}), ..., (\mathbf{x}^{(m)}, \mathbf{y}^{(m)}))\}$



$\Longrightarrow$

x_1 pressão arterial

x_2 temperatura

x_3 peso

x_4 sente fadiga

x_5 tem tosse

$\vdots$

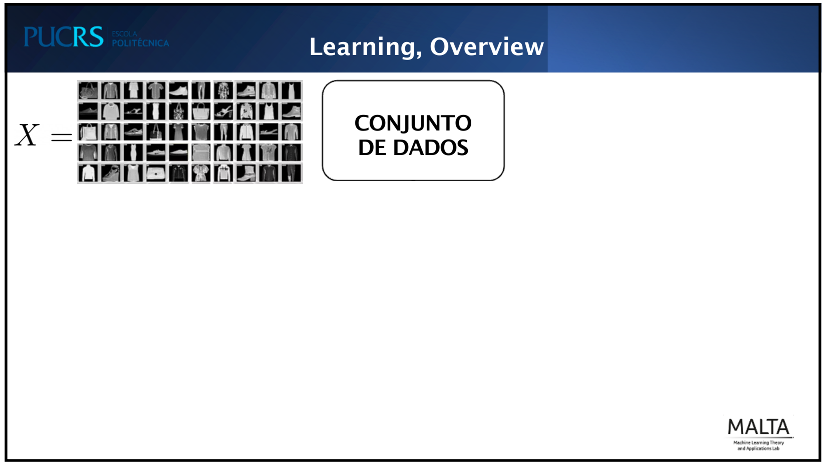
x_d etc

$\mathbf{x}^{(m)} = (x_1, x_2, ..., x_d) \in \mathbb{R}^d$

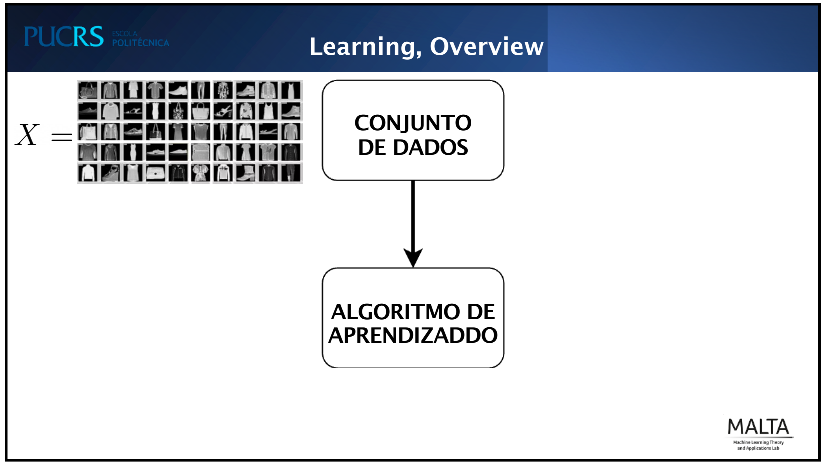
$\mathbf{y}^{(m)} \in \{0, 1\}$

MALTA Machine Learning Theory and Applications Lab

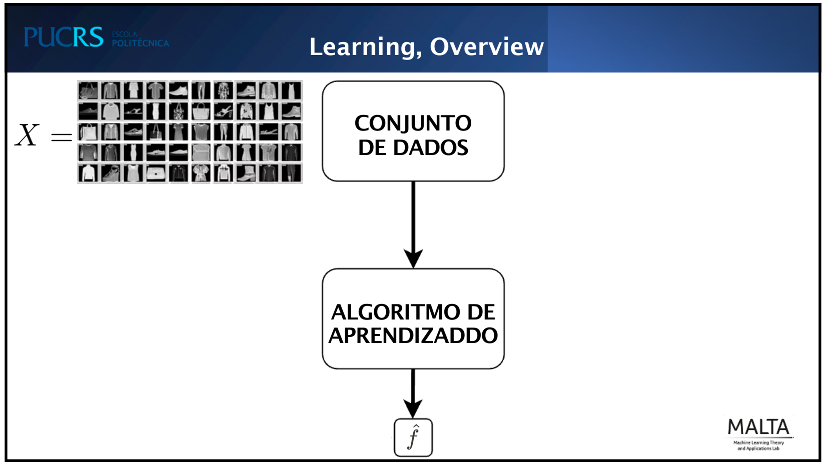
28


PUCRS
Learning, Overview
$X =$
CONJUNTO DE DADOS
MALTA

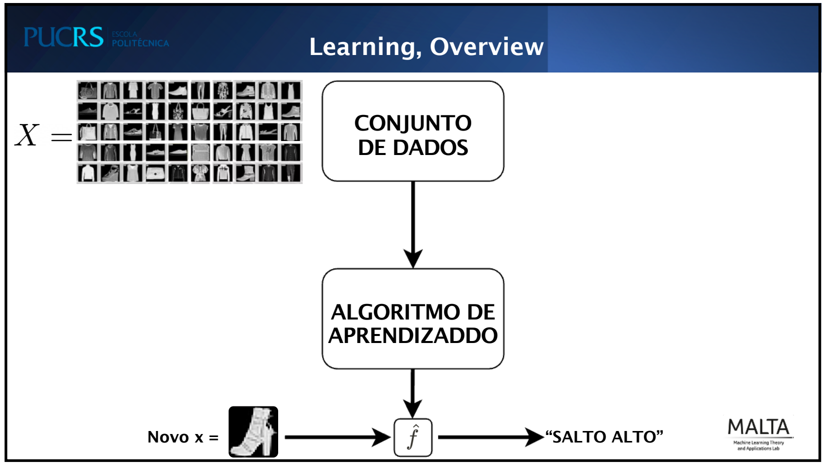
29


PUCRS
Learning, Overview
$X =$
CONJUNTO DE DADOS
ALGORITMO DE APRENDIZADDO
MALTA

30


PUCRS
Learning, Overview
$X =$
CONJUNTO DE DADOS
ALGORITMO DE APRENDIZADDO
$\hat{f}$
MALTA

31


PUCRS
Learning, Overview
$X =$
CONJUNTO DE DADOS
ALGORITMO DE APRENDIZADDO
Novo x =
$\hat{f}$
"SALTO ALTO"
MALTA

32

APRENDIZADO DE MÁQUINA

O QUE É *LEARNING*?

> PRIMEIRO ALG DE LEARNING

PROTOCOLOS DE TREINAM.

OTIMIZAÇÃO

REDES NEURAIS

BACKPROP + PRATICA

CNN + PRATICA

33

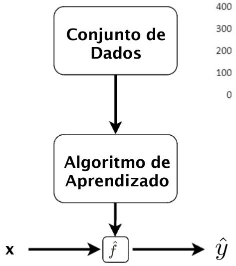
"*Machine Learning is not magic; it cannot get something **from nothing***"
- Domingos

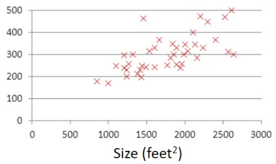
34

PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência





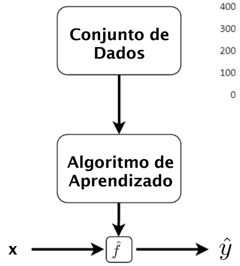
MALTA

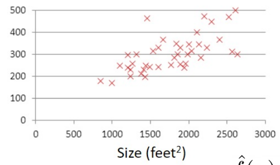
35

PUCRS ESCOLA POLITÉCNICA

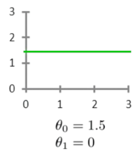
Regressão Linear (Univariada)

Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência



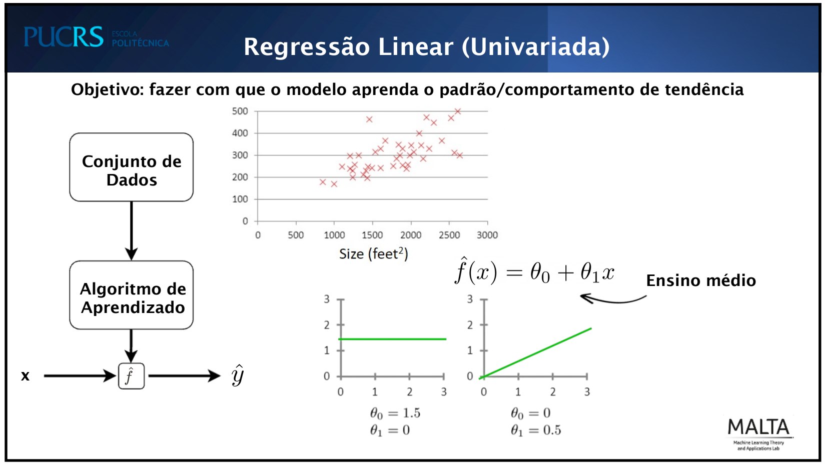


$\hat{f}(x) = \theta_0 + \theta_1 x$ **Ensino médio**

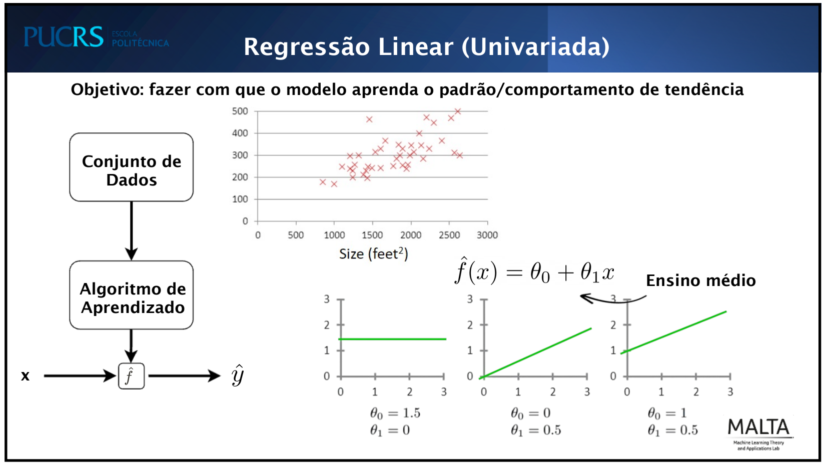


MALTA

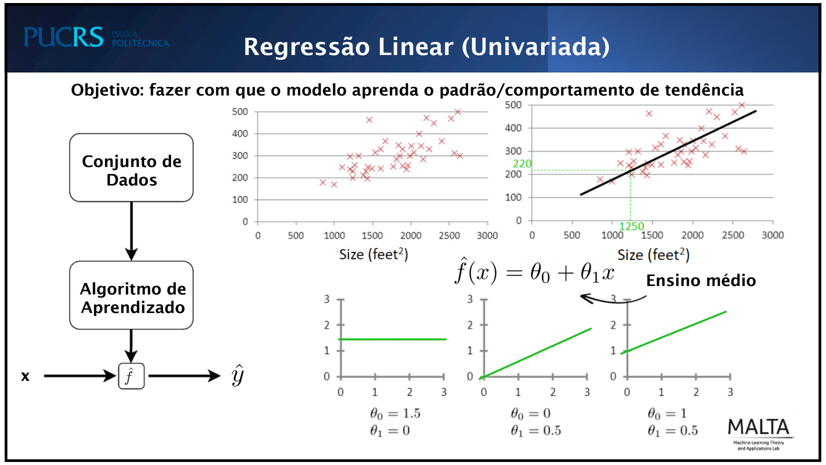
36


Regressão Linear (Univariada)
Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência
Conjunto de Dados
Algoritmo de Aprendizado
Size (feet²)
$\hat{f}(x) = \theta_0 + \theta_1 x$
Ensino médio
$\theta_0 = 1.5$
$\theta_1 = 0$
$\theta_0 = 0$
$\theta_1 = 0.5$
MALTA

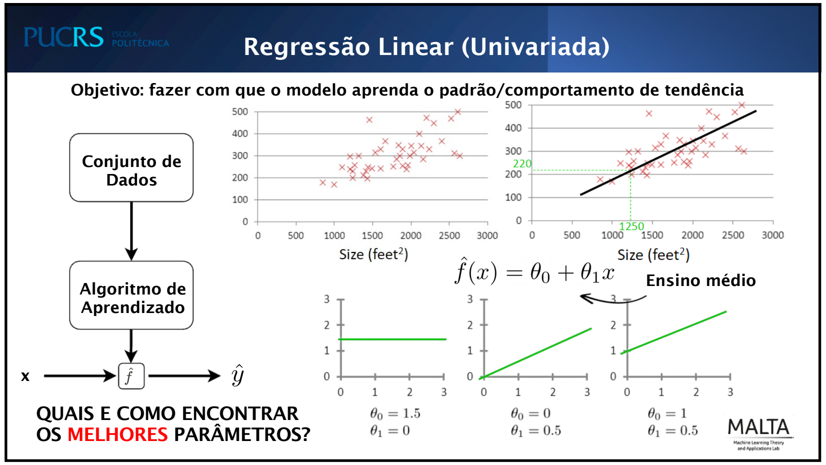
37


Regressão Linear (Univariada)
Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência
Conjunto de Dados
Algoritmo de Aprendizado
Size (feet²)
$\hat{f}(x) = \theta_0 + \theta_1 x$
Ensino médio
$\theta_0 = 1.5$
$\theta_1 = 0$
$\theta_0 = 0$
$\theta_1 = 0.5$
$\theta_0 = 1$
$\theta_1 = 0.5$
MALTA

38

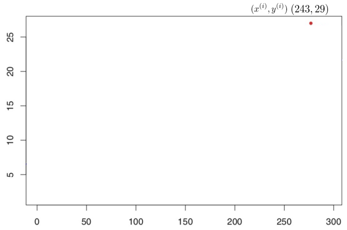

Regressão Linear (Univariada)
Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência
Conjunto de Dados
Algoritmo de Aprendizado
Size (feet²)
220
1250
$\hat{f}(x) = \theta_0 + \theta_1 x$
Ensino médio
$\theta_0 = 1.5$
$\theta_1 = 0$
$\theta_0 = 0$
$\theta_1 = 0.5$
$\theta_0 = 1$
$\theta_1 = 0.5$
MALTA

39


Regressão Linear (Univariada)
Objetivo: fazer com que o modelo aprenda o padrão/comportamento de tendência
Conjunto de Dados
Algoritmo de Aprendizado
Size (feet²)
220
1250
$\hat{f}(x) = \theta_0 + \theta_1 x$
Ensino médio
QUAIS E COMO ENCONTRAR
OS MELHORES PARÂMETROS?
$\theta_0 = 1.5$
$\theta_1 = 0$
$\theta_0 = 0$
$\theta_1 = 0.5$
$\theta_0 = 1$
$\theta_1 = 0.5$
MALTA

40

Regressão Linear (Univariada)

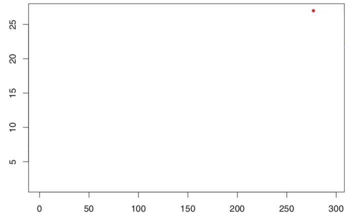


MALTA

41

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

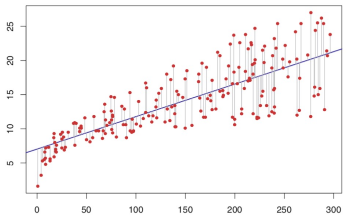


MALTA

42

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$



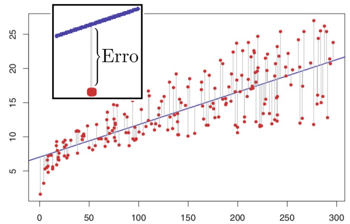
MALTA

43

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)



AVALIANDO OS PARÂMETROS

MALTA

44

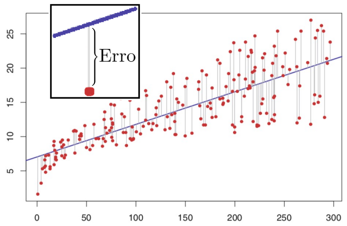
PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\underbrace{\hat{f}\left(x^{(i)}\right) - y^{(i)}}_{\text{Erro}}$$



AVALIANDO OS PARÂMETROS

MALTA

45

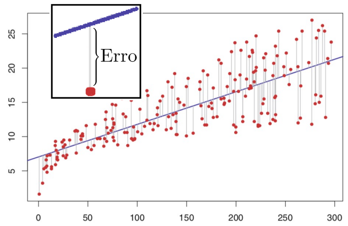
PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\sum_{i=1} \underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)}_{\text{Erro}}$$



AVALIANDO OS PARÂMETROS

MALTA

46

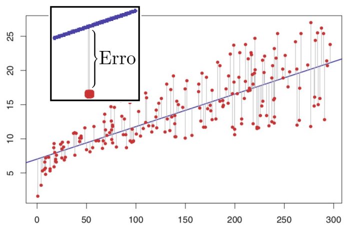
PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\sum_{i=1}^{N} \underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)}_{\text{Erro}}$$



AVALIANDO OS PARÂMETROS

MALTA

47

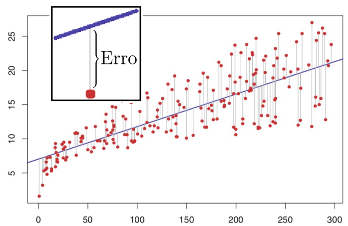
PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\frac{1}{N}\sum_{i=1}^{N} \underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)}_{\text{Erro}}$$



AVALIANDO OS PARÂMETROS

MALTA

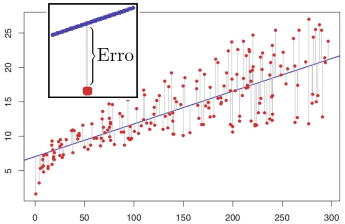
48

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\frac{1}{N}\sum_{i=1}^{N}\underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)^2}_{\text{Erro}}$$



AVALIANDO OS PARÂMETROS

MALTA

49

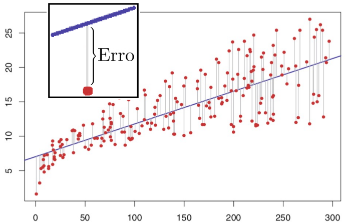
Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\frac{1}{N}\sum_{i=1}^{N}\underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)^2}_{\text{Erro}}$$

$$\underbrace{\theta_0 + \theta_1 x^{(i)}}_{\text{nosso modelo}} - y^{(i)}$$



AVALIANDO OS PARÂMETROS

MALTA

50

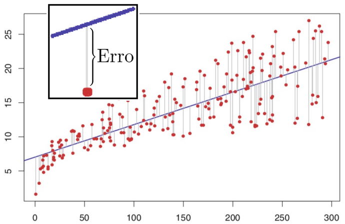
Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\frac{1}{N}\sum_{i=1}^{N}\underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)^2}_{\text{Erro}}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\left(\underbrace{\theta_0 + \theta_1 x^{(i)}}_{\text{nosso modelo}} - y^{(i)}\right)^2$$



AVALIANDO OS PARÂMETROS

MALTA

51

Regressão Linear (Univariada)

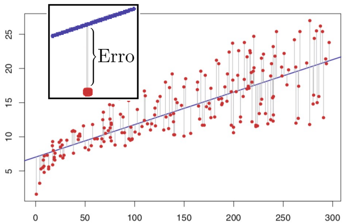
Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall(x^{(i)}, y^{(i)})$

CALCULAR O CUSTO (resíduo/erro/perda)

$$\frac{1}{N}\sum_{i=1}^{N}\underbrace{\left(\hat{f}\left(x^{(i)}\right) - y^{(i)}\right)^2}_{\text{Erro}}$$

$$\frac{1}{2N}\sum_{i=1}^{N}\left(\underbrace{\theta_0 + \theta_1 x^{(i)}}_{\text{nosso modelo}} - y^{(i)}\right)^2$$

$$\underbrace{J(\theta_0, \theta_1)}_{\text{MSE Loss}} = \frac{1}{2N}\sum_{i=1}^{N}\left(\left(\theta_o + \theta_1 x^{(i)}\right) - y^{(i)}\right)^2$$



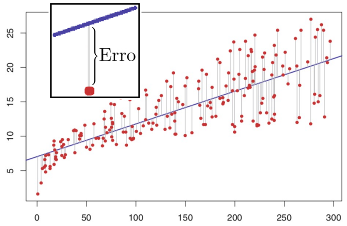
AVALIANDO OS PARÂMETROS

MALTA

52

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$



MALTA

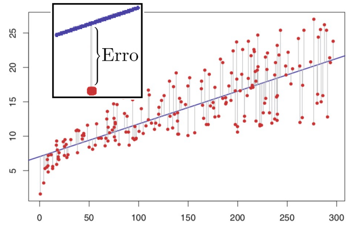
53

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

$$\hat{f}(x) = \theta_0 + \theta_1 x$$



MALTA

54

Regressão Linear (Univariada)

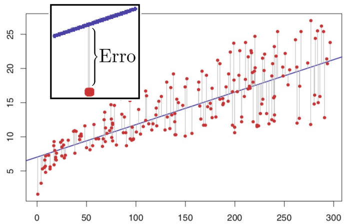
Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Parâmetros

$$\theta_0, \theta_1$$



MALTA

55

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

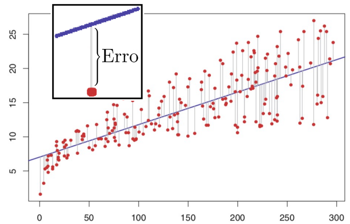
$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Parâmetros

$$\theta_0, \theta_1$$

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$



MALTA

56

PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Parâmetros

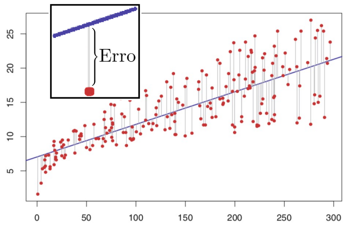
$$\theta_0, \theta_1$$

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$



MALTA

57

PUCRS ESCOLA POLITÉCNICA

Regressão Linear (Univariada)

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

Parâmetros

$$\theta_0, \theta_1$$

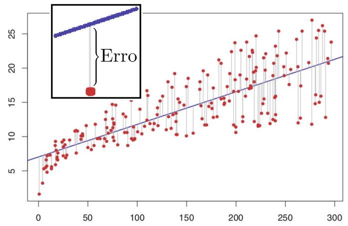
Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO?



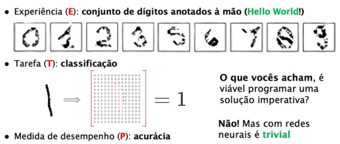
MALTA

58

PUCRS ESCOLA POLITÉCNICA

RECAPITULANDO...

Learning: melhora o desempenho de T conforme a experiência E

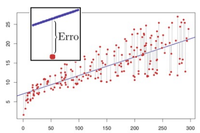


Função de Custo avalia os parâmetros e, precisa ser diferenciável

Ideia: encontrar θ_0, θ_1 de maneira que $\hat{f}(x)$ seja o mais perto de y, $\forall (x^{(i)}, y^{(i)})$

Modelo

$$\hat{f}(x) = \theta_0 + \theta_1 x$$



Parâmetros

$$\theta_0, \theta_1$$

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

MALTA

59

APRENDIZADO DE MÁQUINA

- O QUE É *LEARNING*?
- PRIMEIRO ALG DE LEARNING
- > PROTOCOLOS DE TREINAM.
- OTIMIZAÇÃO
- REDES NEURAIS
- BACKPROP + PRATICA
- CNN + PRATICA

60

PUCRS ESCOLA POLITÉCNICA

Treinamento

Para treinar um modelo, é preciso minimizar a ***loss***, que afeta diretamente o desempenho do modelo no conjunto de **treino**

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

MALTA Machine Learning Theory and Applications Lab

61

PUCRS ESCOLA POLITÉCNICA

Treinamento

Protocolo de Treinamento

Full Dataset

Training Set | **Testing Set**

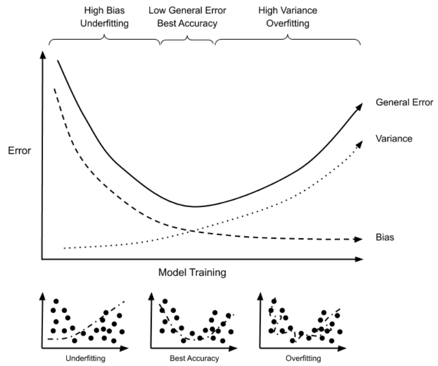
MALTA Machine Learning Theory and Applications Lab

62

PUCRS ESCOLA POLITÉCNICA

Treinamento

Essência de
LEARNING FROM DATA



MALTA Machine Learning Theory and Applications Lab

63

PUCRS ESCOLA POLITÉCNICA

Treinamento

Para treinar um modelo, é preciso minimizar a *loss*, que afeta diretamente o desempenho do modelo no conjunto de **treino**.

Funções de Custo (mais comuns)

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2 \qquad \text{Loss} = - \sum_{i=1}^{\text{output size}} y_i \cdot \log \hat{y}_i$$

MALTA Machine Learning Theory and Applications Lab

64

Treinamento

Para treinar um modelo, é preciso minimizar a ***loss***, que afeta diretamente o desempenho do preditor no conjunto de treino.

Dividir em Treino e Teste

Objetivo de aprendizado de máquina: **generalização**

MALTA

65

APRENDIZADO DE MÁQUINA

O QUE É *LEARNING*?

PRIMEIRO ALG DE LEARNING

PROTOCOLOS DE TREINAM.

> OTIMIZAÇÃO

REDES NEURAIS

BACKPROP + PRATICA

CNN + PRATICA

66

"The connections between neurons are called ***synapses***. The strength of the synaptic connection dictates how much electrical excitation passes from one neuron to another. ***By changing the strength of synaptic connections, animals learn***"

- Bliss & Terge, 1973

67

Gradiente Descendente

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO?

MALTA

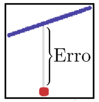
68

PUCRS ESCOLA POLITÉCNICA

Gradiente Descendente

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)}\right)^2$$
$$= \frac{1}{2N}\sum_{i=1}^{N}\underbrace{\left(\hat{y}^{(i)} - y^{(i)}\right)}_{\text{Erro}}^2$$



Objetivo:

$$\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO?

MALTA

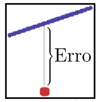
69

PUCRS ESCOLA POLITÉCNICA

Gradiente Descendente

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)}\right)^2$$
$$= \frac{1}{2N}\sum_{i=1}^{N}\underbrace{\left(\hat{y}^{(i)} - y^{(i)}\right)}_{\text{Erro}}^2$$



Objetivo:

$$\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO? Gradiente Descendente

MALTA

70

PUCRS ESCOLA POLITÉCNICA

Gradiente Descendente

Função de custo

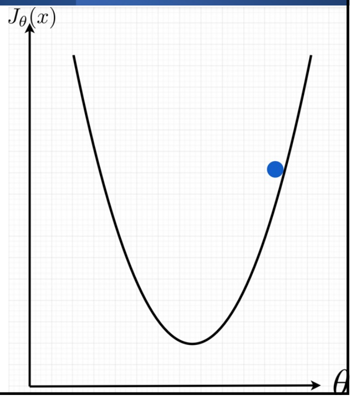
$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)}\right)^2$$

Objetivo:

$$\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO? Gradiente Descendente

1. Iniciar os pesos aleatoriamente



71

PUCRS ESCOLA POLITÉCNICA

Gradiente Descendente

Função de custo

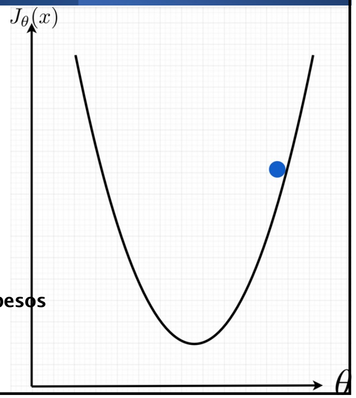
$$J(\theta_0, \theta_1) = \frac{1}{2N}\sum_{i=1}^{N}\left(\left(\theta_0 + \theta_1 x^{(i)}\right) - y^{(i)}\right)^2$$

Objetivo:

$$\min_{\theta_0,\theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO? Gradiente Descendente

1. Iniciar os pesos aleatoriamente
2. Calcular o gradiente da *loss* com relação aos pesos



72

Gradiente Descendente

Função de custo

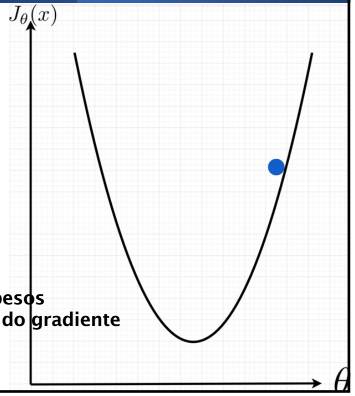
$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO? Gradiente Descendente

1. Iniciar os pesos aleatoriamente
2. Calcular o gradiente da *loss* com relação aos pesos
3. Atualizar os pesos na direção oposta a fração do gradiente



73

Gradiente Descendente

Função de custo

$$J(\theta_0, \theta_1) = \frac{1}{2N} \sum_{i=1}^{N} \left(\left(\theta_0 + \theta_1 x^{(i)} \right) - y^{(i)} \right)^2$$

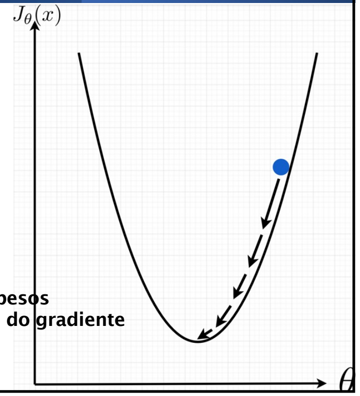
Objetivo:

$$\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$$

COMO FAZER ISSO? Gradiente Descendente

1. Iniciar os pesos aleatoriamente
2. Calcular o gradiente da *loss* com relação aos pesos
3. Atualizar os pesos na direção oposta a fração do gradiente

$$\theta_{n+1} = \theta_n - \alpha \nabla J_\theta$$

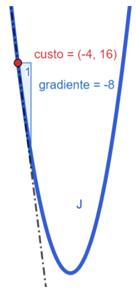


74

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20



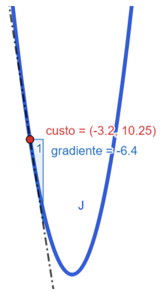
MALTA

75

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56



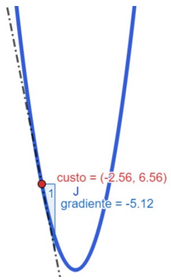
MALTA

76

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04

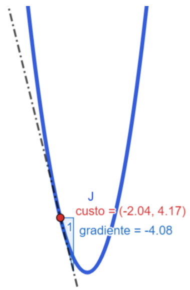


77

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04
3	4.17	-2.04	-4.08	0.1	-1.63

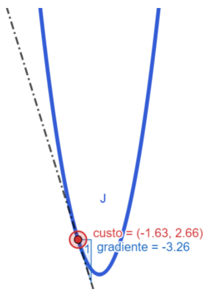


78

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04
3	4.17	-2.04	-4.08	0.1	-1.63
4	2.66	-1.63	-3.26	0.1	-1.30

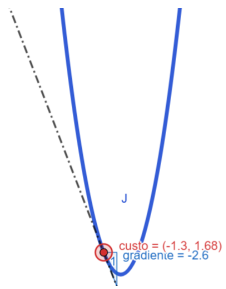


79

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04
3	4.17	-2.04	-4.08	0.1	-1.63
4	2.66	-1.63	-3.26	0.1	-1.30
5	1.68	-1.30	-2.60	0.1	-1.04



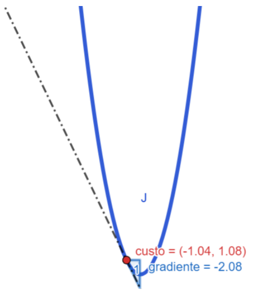
80

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	−4	−8	0.1	−3.20
1	10.25	−3.2	−6.4	0.1	−2.56
2	6.56	−2.56	−5.12	0.1	−2.04
3	4.17	−2.04	−4.08	0.1	−1.63
4	2.66	−1.63	−3.26	0.1	−1.30
5	1.68	−1.30	−2.60	0.1	−1.04
6	1.08	−1.04	−2.08	0.1	−0.83



MALTA

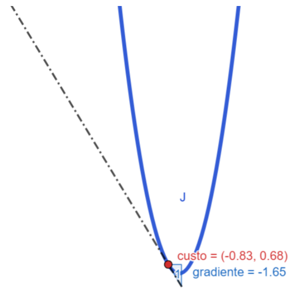
81

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	−4	−8	0.1	−3.20
1	10.25	−3.2	−6.4	0.1	−2.56
2	6.56	−2.56	−5.12	0.1	−2.04
3	4.17	−2.04	−4.08	0.1	−1.63
4	2.66	−1.63	−3.26	0.1	−1.30
5	1.68	−1.30	−2.60	0.1	−1.04
6	1.08	−1.04	−2.08	0.1	−0.83
7	0.68	−0.83	−1.65	0.1	−0.66



MALTA

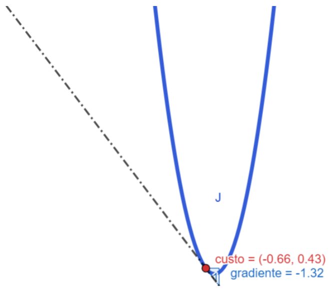
82

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	−4	−8	0.1	−3.20
1	10.25	−3.2	−6.4	0.1	−2.56
2	6.56	−2.56	−5.12	0.1	−2.04
3	4.17	−2.04	−4.08	0.1	−1.63
4	2.66	−1.63	−3.26	0.1	−1.30
5	1.68	−1.30	−2.60	0.1	−1.04
6	1.08	−1.04	−2.08	0.1	−0.83
7	0.68	−0.83	−1.65	0.1	−0.66
8	0.43	−0.66	−1.32	0.1	−0.53



MALTA

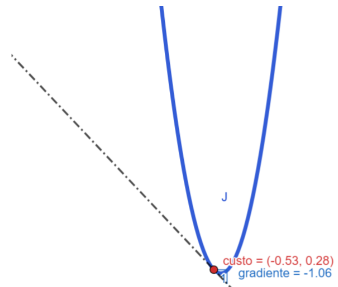
83

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	−4	−8	0.1	−3.20
1	10.25	−3.2	−6.4	0.1	−2.56
2	6.56	−2.56	−5.12	0.1	−2.04
3	4.17	−2.04	−4.08	0.1	−1.63
4	2.66	−1.63	−3.26	0.1	−1.30
5	1.68	−1.30	−2.60	0.1	−1.04
6	1.08	−1.04	−2.08	0.1	−0.83
7	0.68	−0.83	−1.65	0.1	−0.66
8	0.43	−0.66	−1.32	0.1	−0.53
9	0.28	−0.53	−1.06	0.1	−0.42



MALTA

84

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04
3	4.17	-2.04	-4.08	0.1	-1.63
4	2.66	-1.63	-3.26	0.1	-1.30
5	1.68	-1.30	-2.60	0.1	-1.04
6	1.08	-1.04	-2.08	0.1	-0.83
7	0.68	-0.83	-1.65	0.1	-0.66
8	0.43	-0.66	-1.32	0.1	-0.53
9	0.28	-0.53	-1.06	0.1	-0.42
⋮	⋮	⋮	⋮	⋮	⋮

MALTA

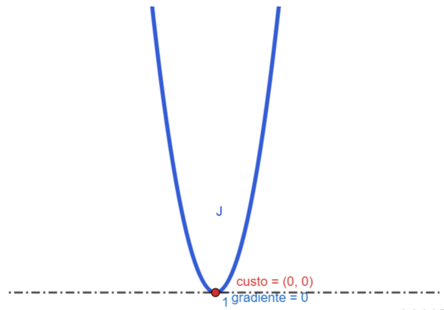
85

PUCRS ESCOLA POLITÉCNICA

Depurando o Gradiente Descendente

$$w_{i+1} = w_i - \alpha \frac{dJ}{dw_i}$$

i	$custo_i$	w_i	$dJ(w_i)$	α	w_{i+1}
0	16.00	-4	-8	0.1	-3.20
1	10.25	-3.2	-6.4	0.1	-2.56
2	6.56	-2.56	-5.12	0.1	-2.04
3	4.17	-2.04	-4.08	0.1	-1.63
4	2.66	-1.63	-3.26	0.1	-1.30
5	1.68	-1.30	-2.60	0.1	-1.04
6	1.08	-1.04	-2.08	0.1	-0.83
7	0.68	-0.83	-1.65	0.1	-0.66
8	0.43	-0.66	-1.32	0.1	-0.53
9	0.28	-0.53	-1.06	0.1	-0.42
⋮	⋮	⋮	⋮	⋮	⋮
n	0	w_{n-1}	0	0.1	0



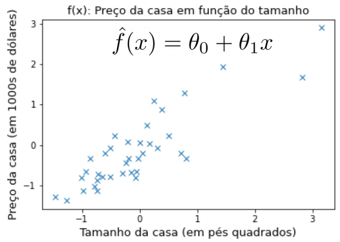
MALTA

86

PUCRS ESCOLA POLITÉCNICA

Exemplo Regressão Linear (Univariada)

Tarefa de regressão: predizer o preço dado o tamanho da casa



$$\theta_{n+1} = \theta_n - \alpha \nabla J_\theta$$

MALTA

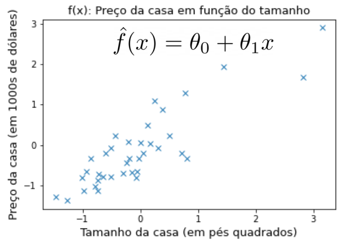
87

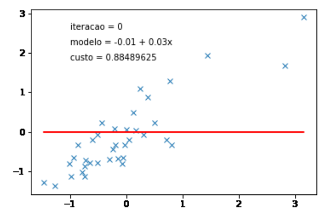
PUCRS ESCOLA POLITÉCNICA

Exemplo Regressão Linear (Univariada)

Tarefa de regressão: predizer o preço dado o tamanho da casa

Objetivo: $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$





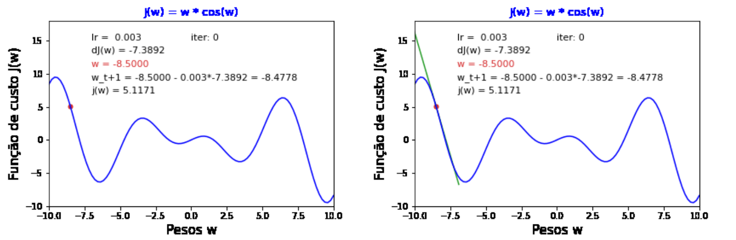
$$\theta_{n+1} = \theta_n - \alpha \nabla J_\theta$$

MALTA

88

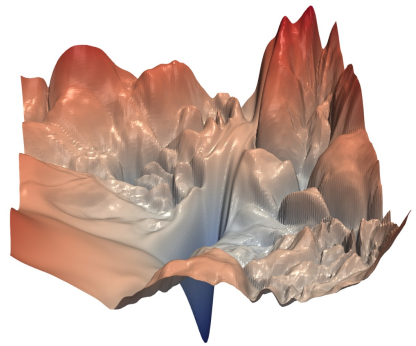
Otimização via Gradiente Descendente

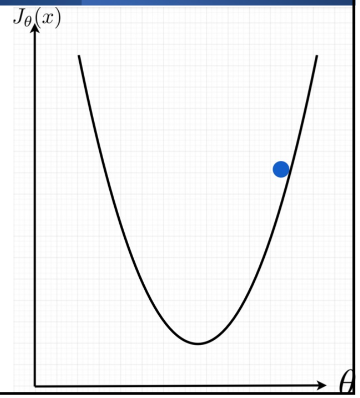
Loss landscape, "Paisagem do custo"



89

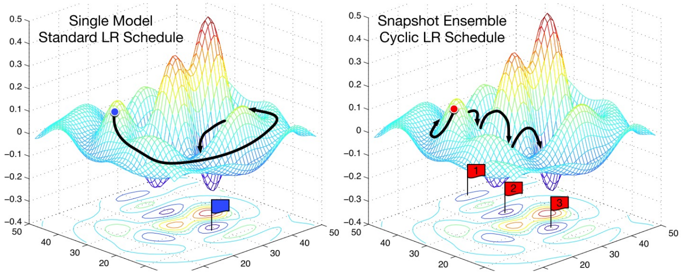
Depurando o Gradiente Descendente





90

Depurando o Gradiente Descendente



91

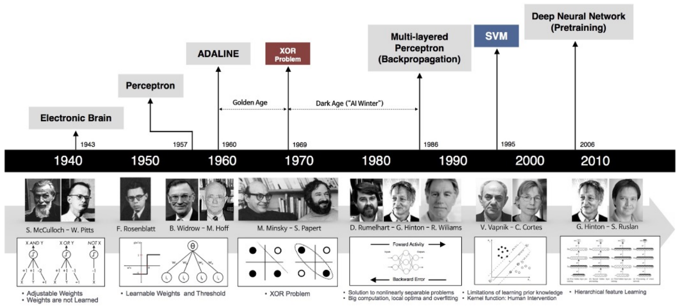
APRENDIZADO DE MÁQUINA

- O QUE É *LEARNING?*
- PRIMEIRO ALG DE LEARNING
- PROTOCOLOS DE TREINAM.
- OTIMIZAÇÃO
- > REDES NEURAIS
- BACKPROP + PRATICA
- CNN + PRATICA

92

PUCRS ESCOLA POLITÉCNICA

Histórico das Redes Neurais



MALTA Machine Learning Theory and Applications Lab

93

PUCRS ESCOLA POLITÉCNICA

Redes Neurais

$$\hat{f}(x) = W_0 + W_1 x$$

$$\hat{f}(x) = \theta_0 + \theta_1 x$$

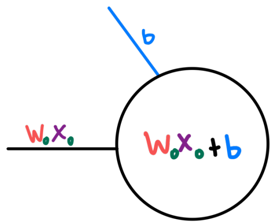
MALTA Machine Learning Theory and Applications Lab

94

PUCRS ESCOLA POLITÉCNICA

Redes Neurais

$$\hat{f}(x) = W_0 + W_1 x$$



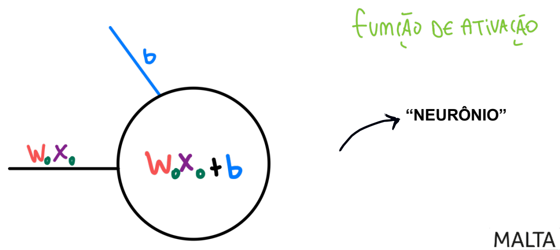
MALTA Machine Learning Theory and Applications Lab

95

PUCRS ESCOLA POLITÉCNICA

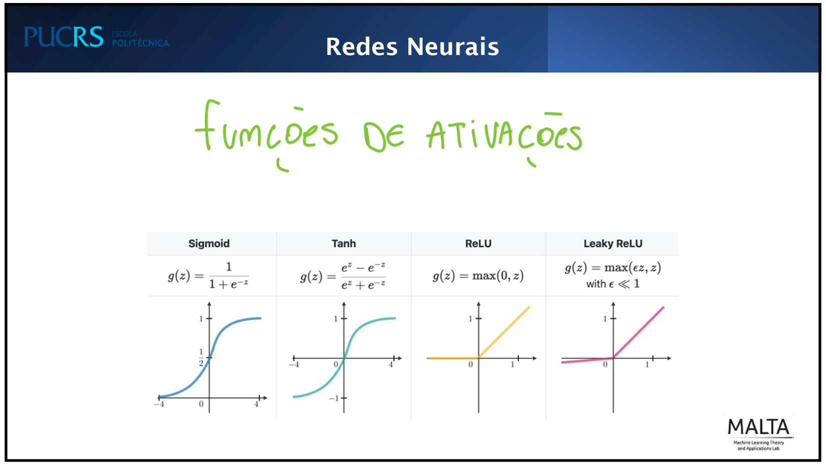
Redes Neurais

$$\hat{f}(x) = W_0 + W_1 x$$

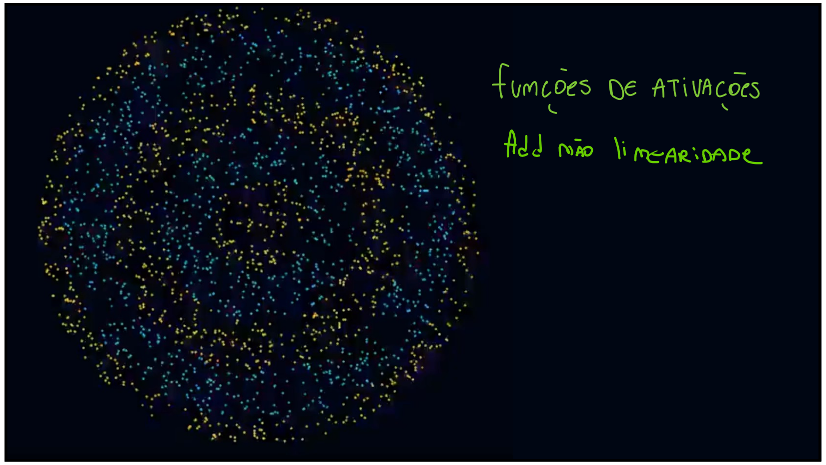


MALTA Machine Learning Theory and Applications Lab

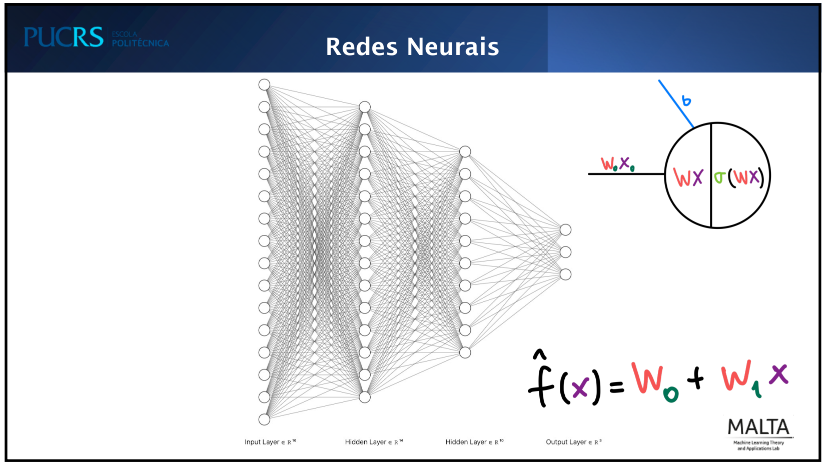
96


PUCRS ESCOLA POLITÉCNICA
Redes Neurais
funções de ativações
Sigmoid
Tanh
ReLU
Leaky ReLU
g(z) = 1 / (1 + e^-z)
g(z) = (e^z − e^-z) / (e^z + e^-z)
g(z) = max(0, z)
g(z) = max(εz, z)
with ε ≪ 1
MALTA

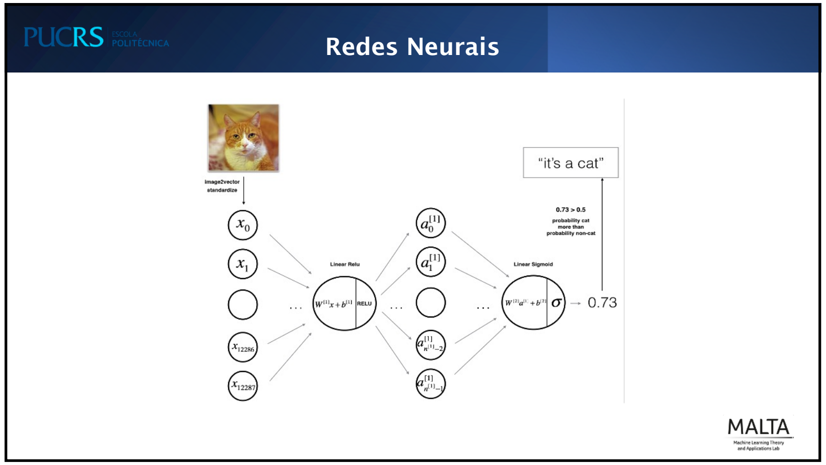
97


funções de ativações
Add não linearidade

98

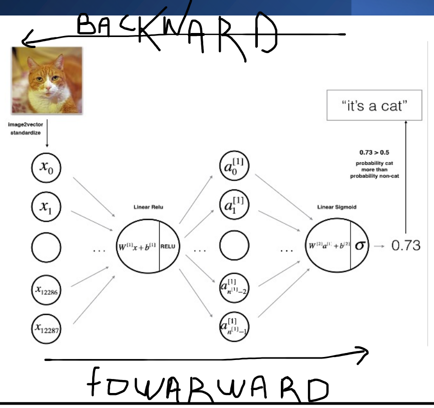

PUCRS ESCOLA POLITÉCNICA
Redes Neurais
b
W0X0
WX
σ(WX)
f̂(x) = W0 + W1 X
MALTA

99


PUCRS ESCOLA POLITÉCNICA
Redes Neurais
"it's a cat"
0.73
MALTA

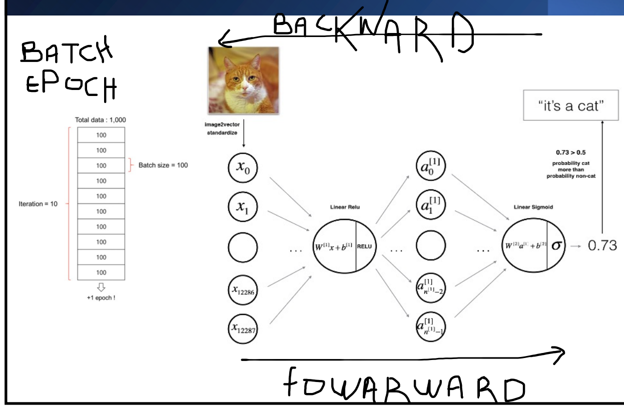
100

Redes Neurais



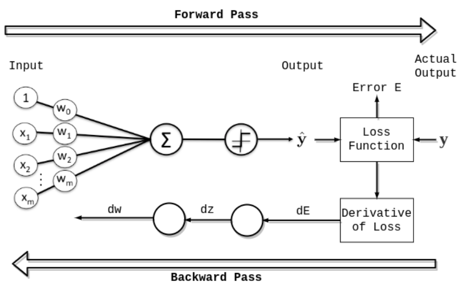
101

Redes Neurais



102

RECAPITULANDO...



103

RECAPITULANDO...

NN é uma **composição** de funções lineares e não lineares.

Funções de **ativações** quebram as transformações lineares

Funções de ativações devem ser **diferenciaveis**

Forward: $\hat{y}$
Backward: calcula os **gradientes** da função de custo em relação a cada parâmetro

104

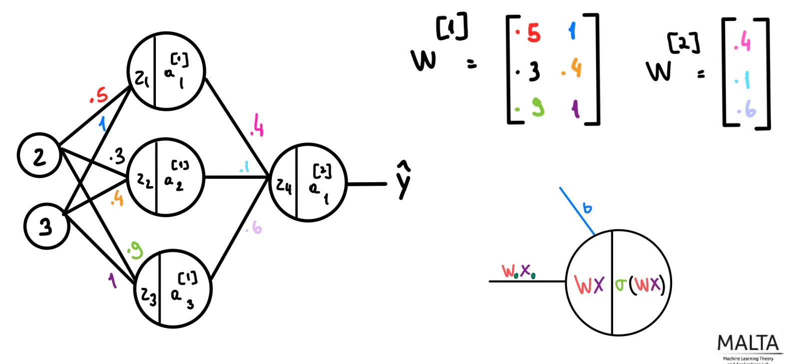
APRENDIZADO DE MÁQUINA

- O QUE É *LEARNING*?
- PRIMEIRO ALG DE LEARNING
- PROTOCOLOS DE TREINAM.
- OTIMIZAÇÃO
- REDES NEURAIS
- > BACKPROP + PRATICA
- CNN + PRATICA

105

PUCRS ESCOLA POLITÉCNICA

Backpropagation

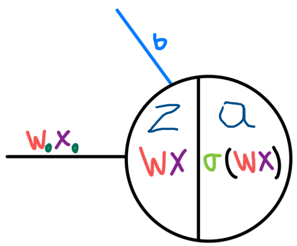


MALTA

106

PUCRS ESCOLA POLITÉCNICA

Redes Neurais

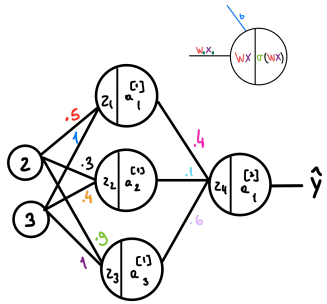


MALTA

107

PUCRS ESCOLA POLITÉCNICA

Redes Neurais



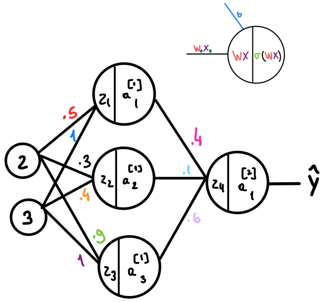
PRIMEIRA CAMADA

$z_1 = 2 \times .5 + 3 \times 1 = 4$, $a_1^{[1]} =$

108

PUCRS ESCOLA POLITÉCNICA

Redes Neurais



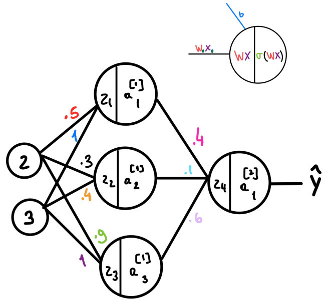
\# PRIMEIRA CAMADA

$z_1 = 2 \times .5 + 3 \times 1 = 4, \quad a_1^{[1]} =$

109

PUCRS ESCOLA POLITÉCNICA

Redes Neurais



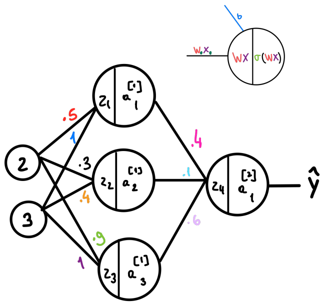
\# PRIMEIRA CAMADA

$z_1 = 2 \times .5 + 3 \times 1 = 4, \quad a_1^{[1]} = \frac{1}{1 + e^{-z_1}} = .9820$

110

PUCRS ESCOLA POLITÉCNICA

Redes Neurais



\# PRIMEIRA CAMADA

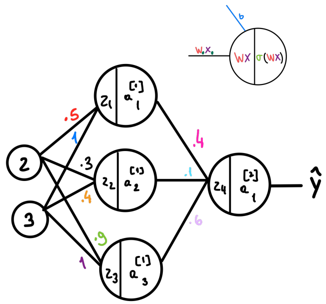
$z_1 = 2 \times .5 + 3 \times 1 = 4, \quad a_1^{[1]} = \frac{1}{1 + e^{-z_1}} = .9820$

$z_2 = 2 \times .3 + 3 \times .4 = 1.8, \quad a_2^{[1]} = \max(0, z_2) = 1.8$

111

PUCRS ESCOLA POLITÉCNICA

Redes Neurais



\# PRIMEIRA CAMADA

$z_1 = 2 \times .5 + 3 \times 1 = 4, \quad a_1^{[1]} = \frac{1}{1 + e^{-z_1}} = .9820$

$z_2 = 2 \times .3 + 3 \times .4 = 1.8, \quad a_2^{[1]} = \max(0, z_2) = 1.8$

$z_3 = 2 \times .9 + 3 \times 1 = 4.8, \quad a_3^{[1]} = \frac{e^{z_3} - e^{-z_3}}{e^{z_3} + e^{-z_3}} = .9999$

\# SEGUNDA CAMADA

$z_4 = a_1^{[1]} \times .4 + a_2^{[1]} \times .1 + a_3^{[1]} \times .6$

$z_4 = .9820 \times .4 + 1.8 \times .1 + .9999 \times .6 = 2.7927$

$a_1^{[2]} = \max(0, z_4) = 2.7927$

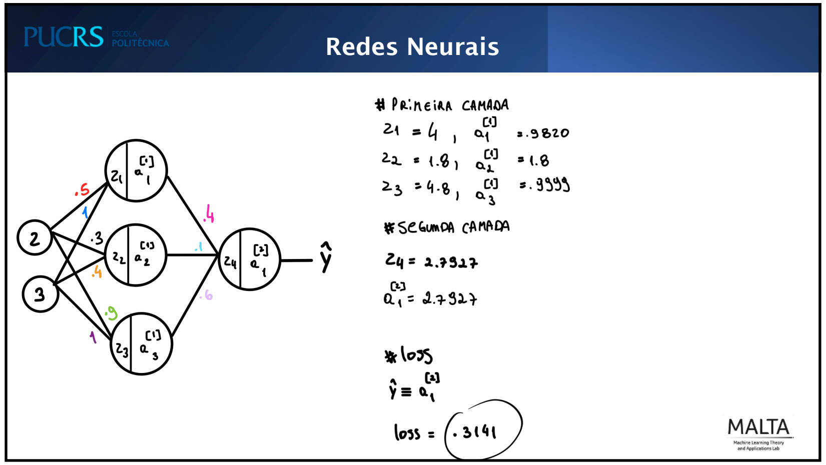
\# loss

$\hat{y} \equiv a_1^{[2]}; \quad x_1 = 2 \quad x_2 = 3 \rightarrow y = 2$

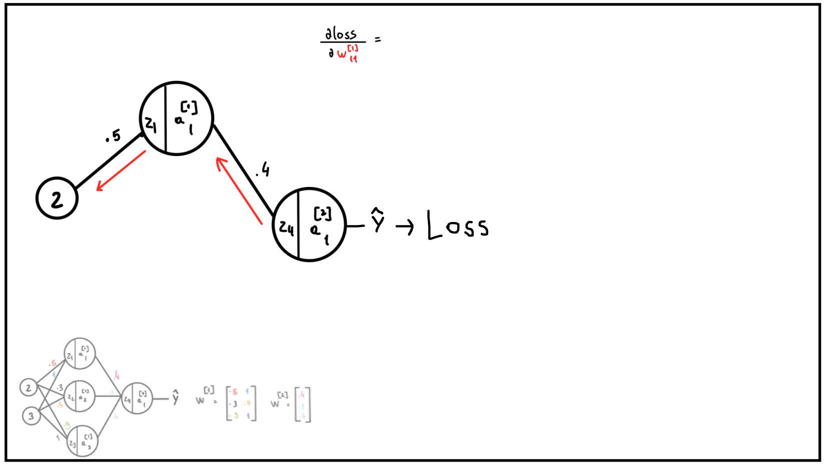
$loss = \frac{1}{2}(a^{[2]} - y)^2 = \frac{1}{2}(2.7927 - 2)^2 = .3141$

MALTA Machine Learning Theory and Applications Lab

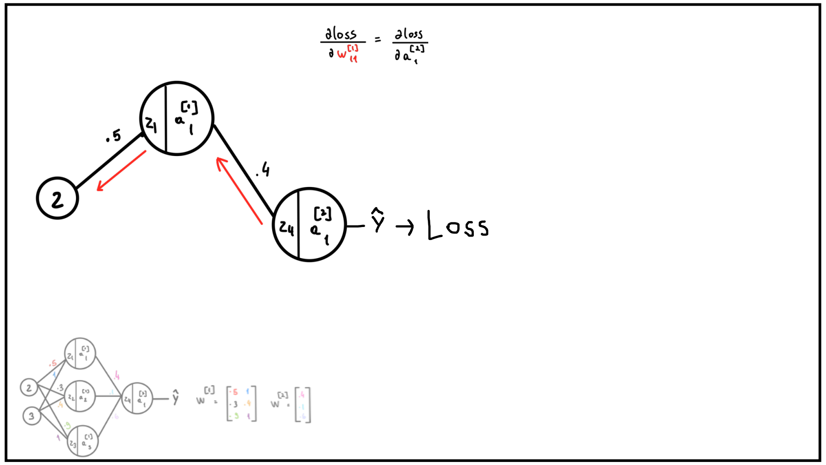
112



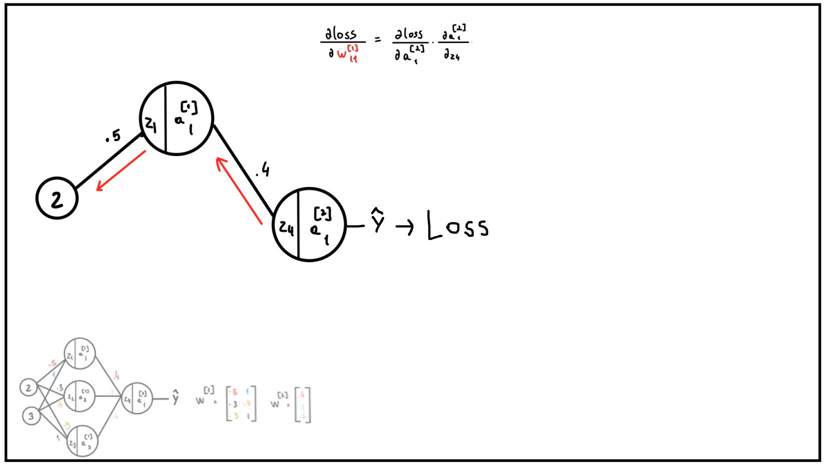
113



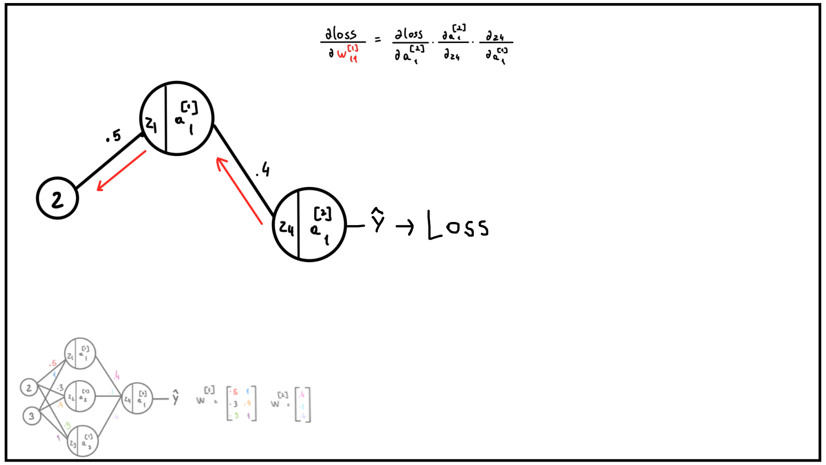
114



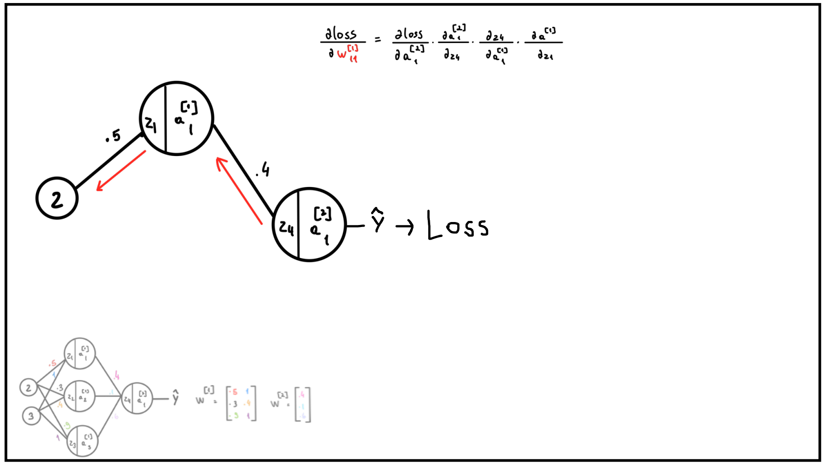
115



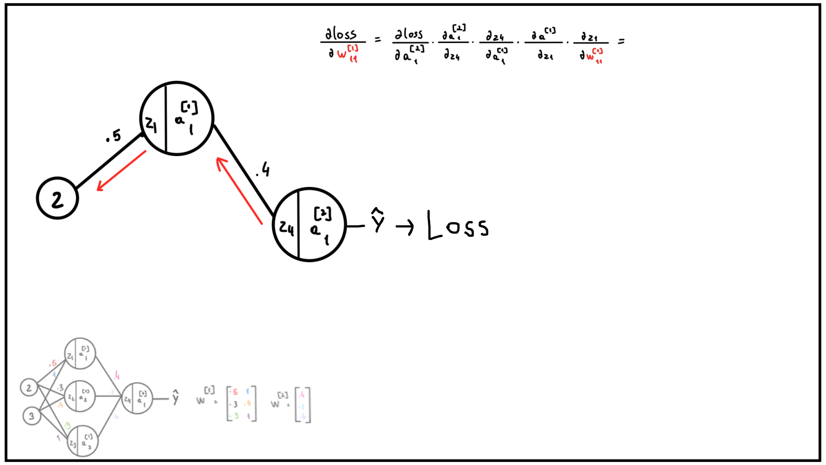
116



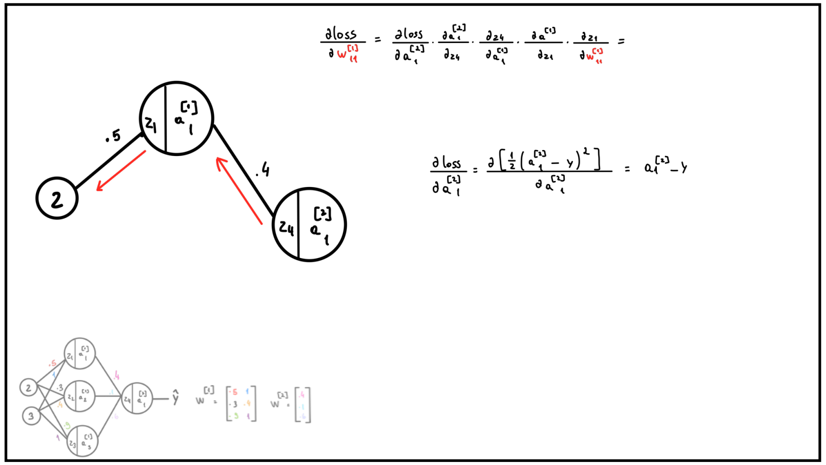
117



118

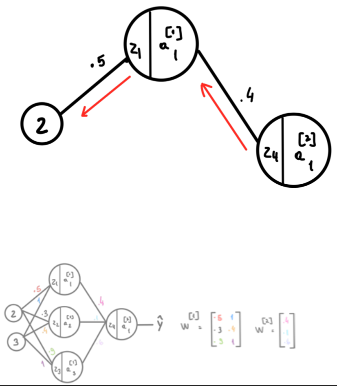


119



120

$$\frac{\partial loss}{\partial w^{[1]}_{11}} = \frac{\partial loss}{\partial a^{[2]}_1}\cdot\frac{\partial a^{[2]}_1}{\partial z_4}\cdot\frac{\partial z_4}{\partial a^{[1]}_1}\cdot\frac{\partial a^{[1]}}{\partial z_1}\cdot\frac{\partial z_1}{\partial w^{[1]}_{11}} =$$

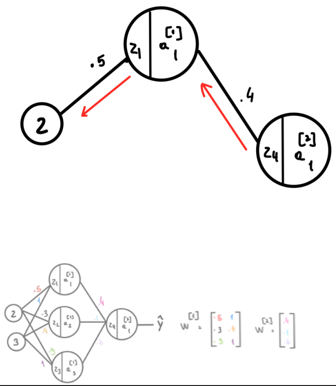


$$\frac{\partial loss}{\partial a^{[2]}_1} = \frac{\partial\left[\frac{1}{2}\left(a^{[2]}_1 - y\right)^2\right]}{\partial a^{[2]}_1} = a^{[2]}_1 - y$$

$$\frac{\partial a^{[2]}_1}{\partial z_4} = \frac{\partial[\max(0,z_4)]}{\partial z_4} = 1(z_4>0)$$

121

$$\frac{\partial loss}{\partial w^{[1]}_{11}} = \frac{\partial loss}{\partial a^{[2]}_1}\cdot\frac{\partial a^{[2]}_1}{\partial z_4}\cdot\frac{\partial z_4}{\partial a^{[1]}_1}\cdot\frac{\partial a^{[1]}}{\partial z_1}\cdot\frac{\partial z_1}{\partial w^{[1]}_{11}} =$$



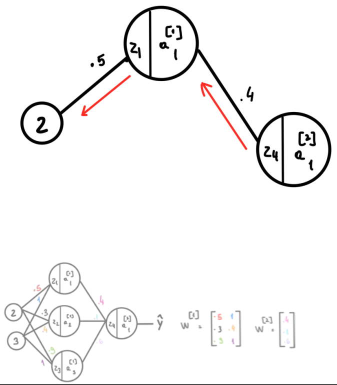
$$\frac{\partial loss}{\partial a^{[2]}_1} = \frac{\partial\left[\frac{1}{2}\left(a^{[2]}_1 - y\right)^2\right]}{\partial a^{[2]}_1} = a^{[2]}_1 - y$$

$$\frac{\partial a^{[2]}_1}{\partial z_4} = \frac{\partial[\max(0,z_4)]}{\partial z_4} = 1(z_4>0)$$

$$\frac{\partial z_4}{\partial a^{[1]}_1} = \frac{\partial[a^{[1]}_1 \times .4 + a^{[1]}_2\times .1 + a^{[1]}_3 \times .6]}{\partial a^{[1]}_1} = .4$$

122

$$\frac{\partial loss}{\partial w^{[1]}_{11}} = \frac{\partial loss}{\partial a^{[2]}_1}\cdot\frac{\partial a^{[2]}_1}{\partial z_4}\cdot\frac{\partial z_4}{\partial a^{[1]}_1}\cdot\frac{\partial a^{[1]}}{\partial z_1}\cdot\frac{\partial z_1}{\partial w^{[1]}_{11}} =$$



$$\frac{\partial loss}{\partial a^{[2]}_1} = \frac{\partial\left[\frac{1}{2}\left(a^{[2]}_1 - y\right)^2\right]}{\partial a^{[2]}_1} = a^{[2]}_1 - y$$

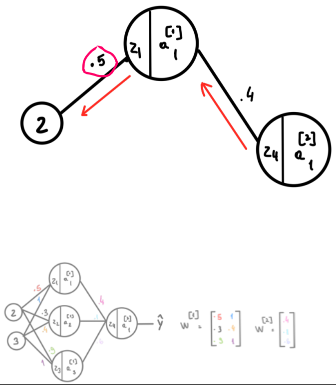
$$\frac{\partial a^{[2]}_1}{\partial z_4} = \frac{\partial[\max(0,z_4)]}{\partial z_4} = 1(z_4>0)$$

$$\frac{\partial z_4}{\partial a^{[1]}_1} = \frac{\partial[a^{[1]}_1 \times .4 + a^{[1]}_2\times .1 + a^{[1]}_3 \times .6]}{\partial a^{[1]}_1} = .4$$

$$\frac{\partial a^{[1]}_1}{\partial z_1} = \frac{\partial\left[\frac{1}{1+e^{-z_1}}\right]}{\partial z_1} = \sigma(z_1)\cdot(1-\sigma(z_1))$$

123

$$\frac{\partial loss}{\partial w^{[1]}_{11}} = \frac{\partial loss}{\partial a^{[2]}_1}\cdot\frac{\partial a^{[2]}_1}{\partial z_4}\cdot\frac{\partial z_4}{\partial a^{[1]}_1}\cdot\frac{\partial a^{[1]}}{\partial z_1}\cdot\frac{\partial z_1}{\partial w^{[1]}_{11}} =$$



$$\frac{\partial loss}{\partial a^{[2]}_1} = \frac{\partial\left[\frac{1}{2}\left(a^{[2]}_1 - y\right)^2\right]}{\partial a^{[2]}_1} = a^{[2]}_1 - y$$

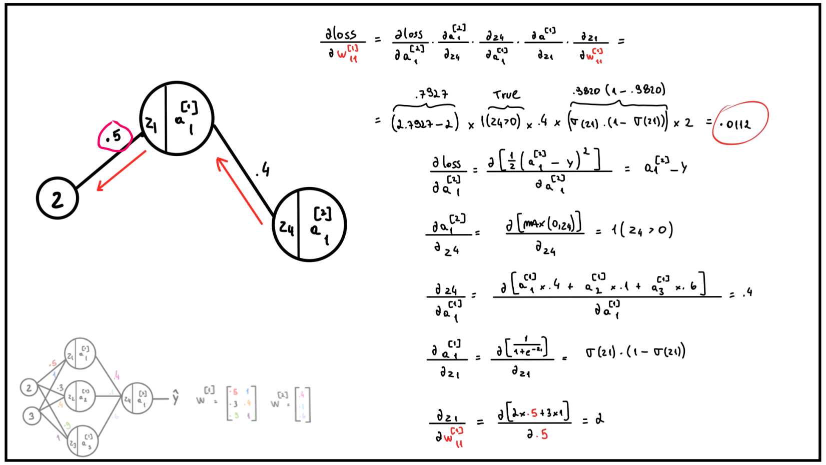
$$\frac{\partial a^{[2]}_1}{\partial z_4} = \frac{\partial[\max(0,z_4)]}{\partial z_4} = 1(z_4>0)$$

$$\frac{\partial z_4}{\partial a^{[1]}_1} = \frac{\partial[a^{[1]}_1 \times .4 + a^{[1]}_2\times .1 + a^{[1]}_3 \times .6]}{\partial a^{[1]}_1} = .4$$

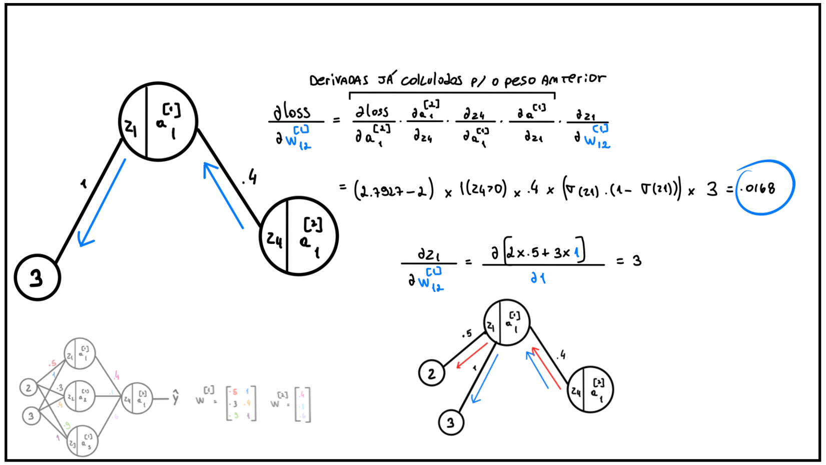
$$\frac{\partial a^{[1]}_1}{\partial z_1} = \frac{\partial\left[\frac{1}{1+e^{-z_1}}\right]}{\partial z_1} = \sigma(z_1)\cdot(1-\sigma(z_1))$$

$$\frac{\partial z_1}{\partial w^{[1]}_{11}} = \frac{\partial[2\times .5 + 3\times 1]}{\partial .5} = 2$$

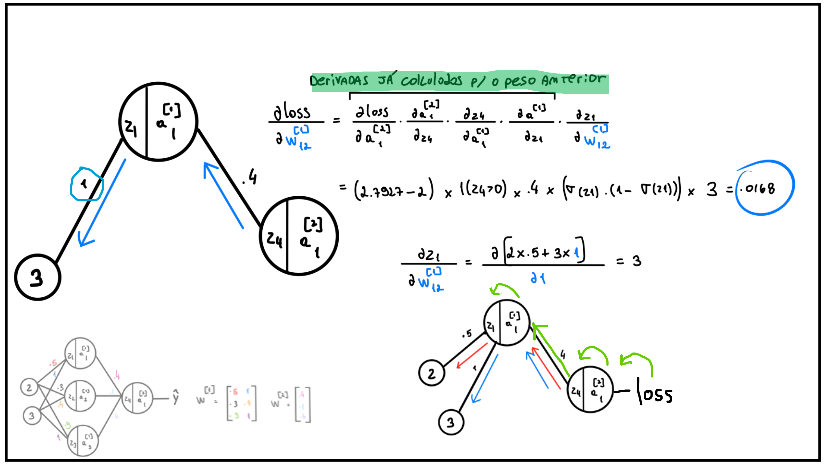
124



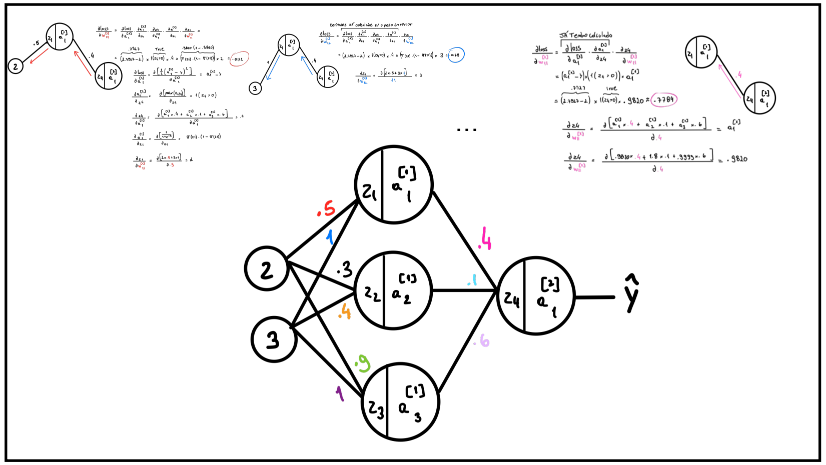
125



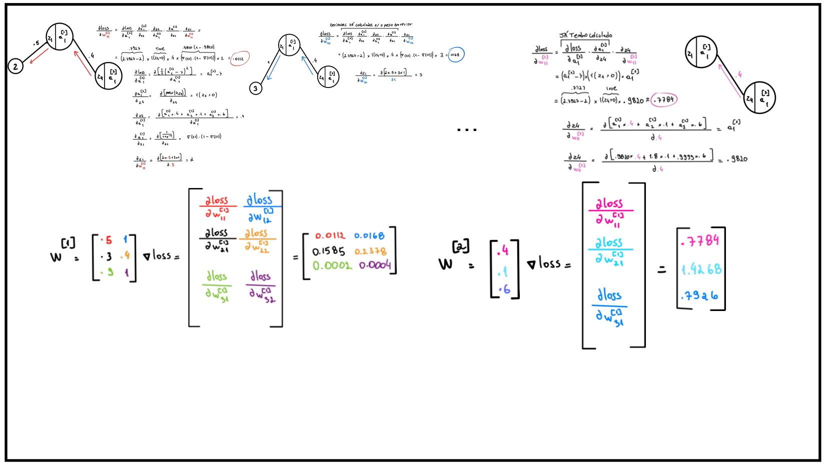
126



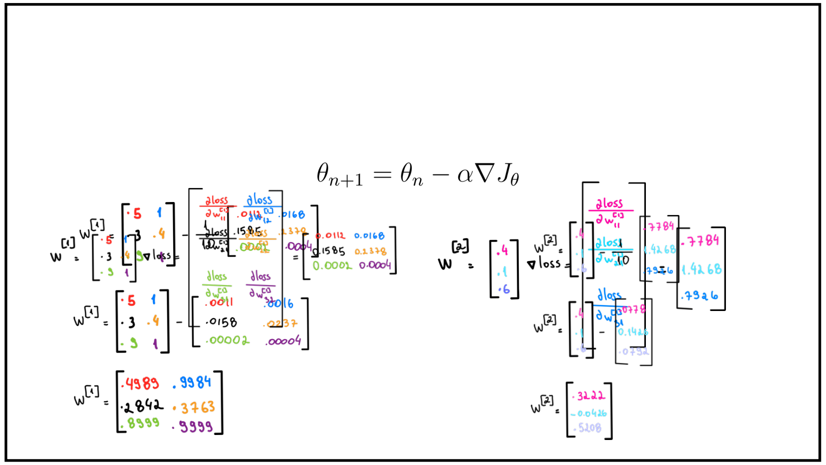
127



128



129

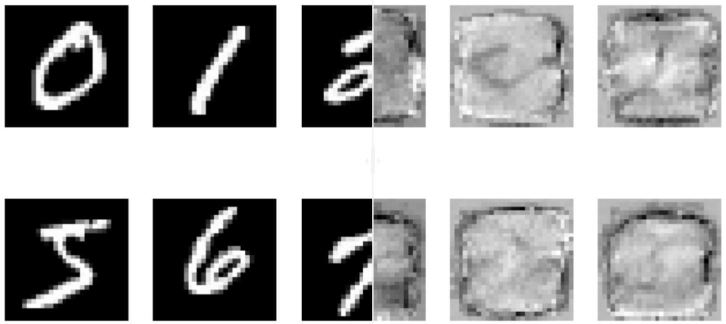


130

PUCRS ESCOLA POLITÉCNICA

PRATICA

shorturl.at/pqJP0



131

APRENDIZADO DE MÁQUINA

O QUE É *LEARNING*?

PRIMEIRO ALG DE LEARNING

PROTOCOLOS DE TREINAM.

OTIMIZAÇÃO

REDES NEURAIS

BACKPROP + PRATICA

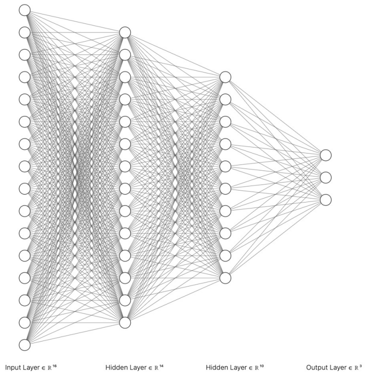
> CNN + PRATICA

132

PUCRS ESCOLA POLITÉCNICA

Convolutional Neural Network

Tarefa envolve images?



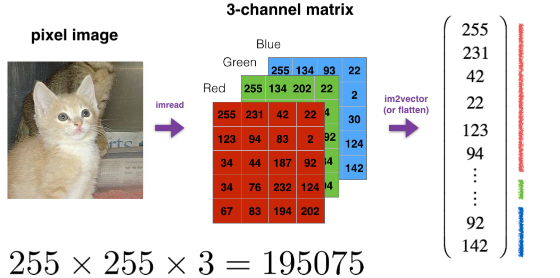
MALTA Machine Learning Theory and Applications Lab

133

PUCRS ESCOLA POLITÉCNICA

Convolutional Neural Network

Tarefa envolve images?



$$255 \times 255 \times 3 = 195075$$

MALTA Machine Learning Theory and Applications Lab

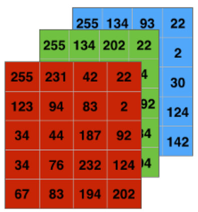
134

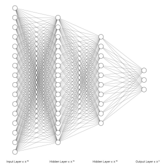
PUCRS ESCOLA POLITÉCNICA

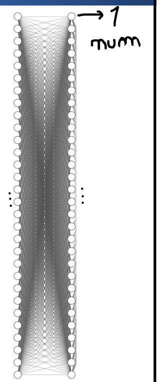
Convolutional Neural Network

QUANTAS OPERAÇÕES SÃO NECESSÁRIAS PARA O PRODUTO DA CAMADA SUBSEQUENTE DA NN MLP RETORNAR UM VETOR DE MESMA DIMENSÃO QUE A IMAGEM DE ENTRADA?

$$255 \times 255 \times 3 = 195075$$





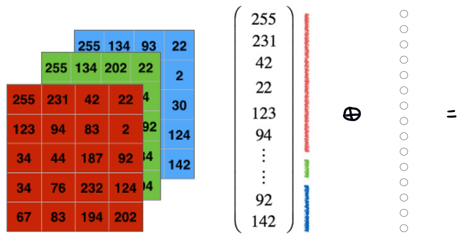


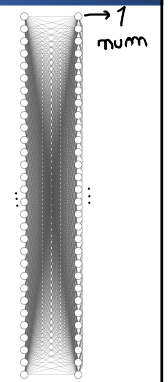
135

PUCRS ESCOLA POLITÉCNICA

Convolutional Neural Network

$$255 \times 255 \times 3 = 195075$$



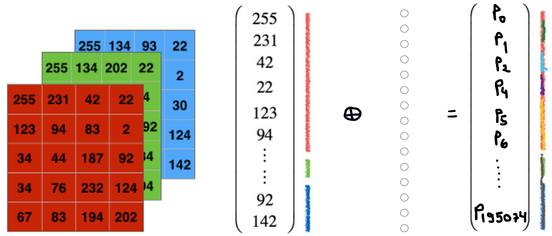


136

PUCRS ESCOLA POLITÉCNICA

Convolutional Neural Network

$$255 \times 255 \times 3 = 195075$$



MALTA

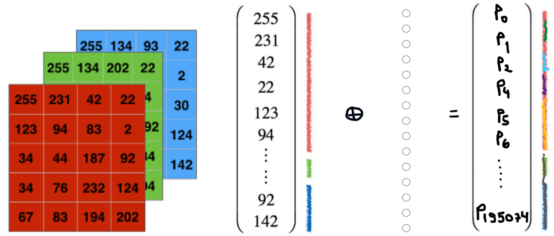
137

PUCRS ESCOLA POLITÉCNICA

Convolutional Neural Network

$$255 \times 255 \times 3 = 195075$$

$$\#\text{Operações} = (1, 195075) \times (195075, 195075) \approx 4 \times 10^{10}$$



MALTA

138

PUCRS ESCOLA POLITÉCNICA

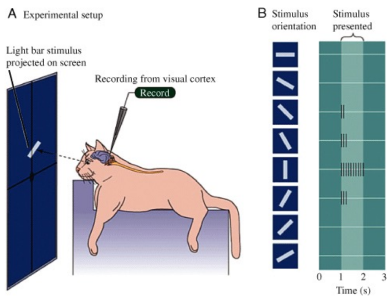
O experimento de Hubel e Wiesel (1959)

J Physiol. 1959 Oct;148(3):574-91. doi: 10.1113/jphysiol.1959.sp006308.

Receptive fields of single neurones in the cat's striate cortex

D H HUBEL, T N WIESEL

PMID: 14403679 PMCID: PMC1363130 DOI: 10.1113/jphysiol.1959.sp006308



https://www.youtube.com/watch?v=iw6nBWo21Zk&t=315s

MALTA

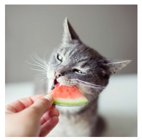
139

PUCRS ESCOLA POLITÉCNICA

O que podemos tirar de insight

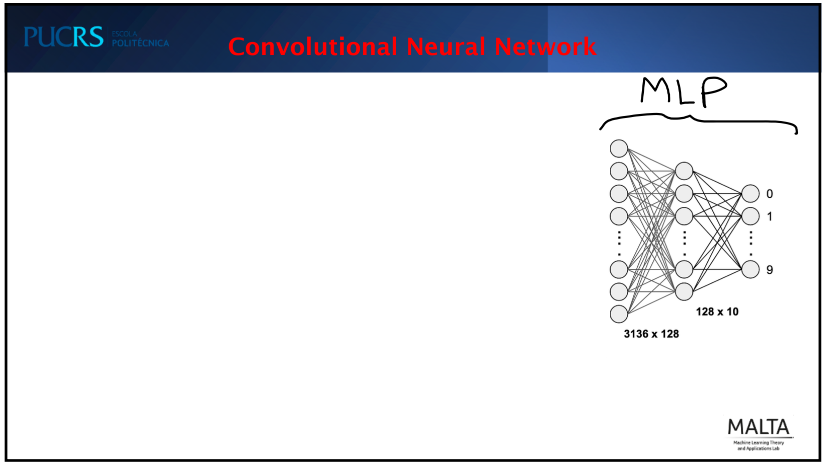
Imagens possuem algumas características interessantes

- Coerência espacial
- Invariância a translação (rotação, escala, ...)

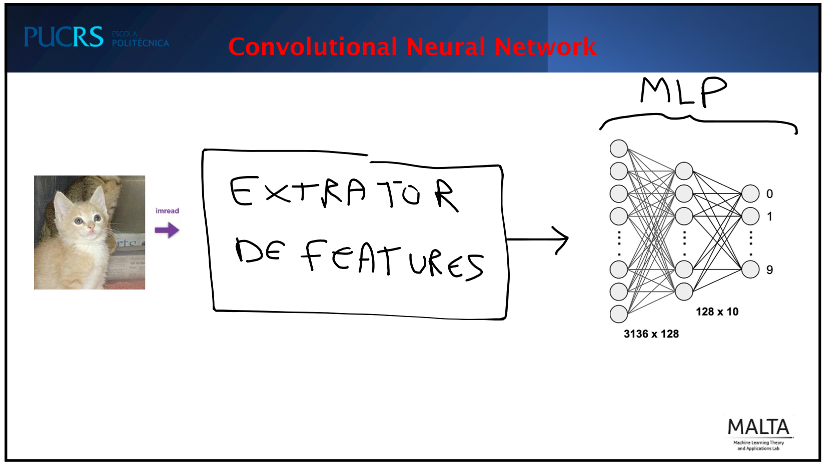


MALTA

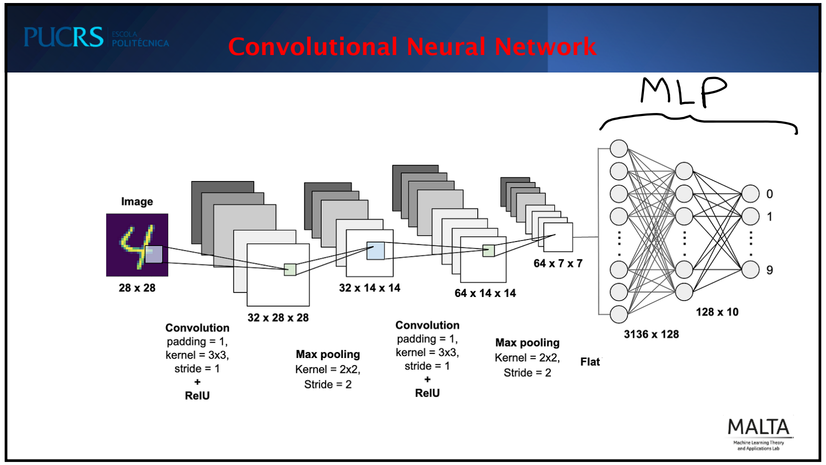
140



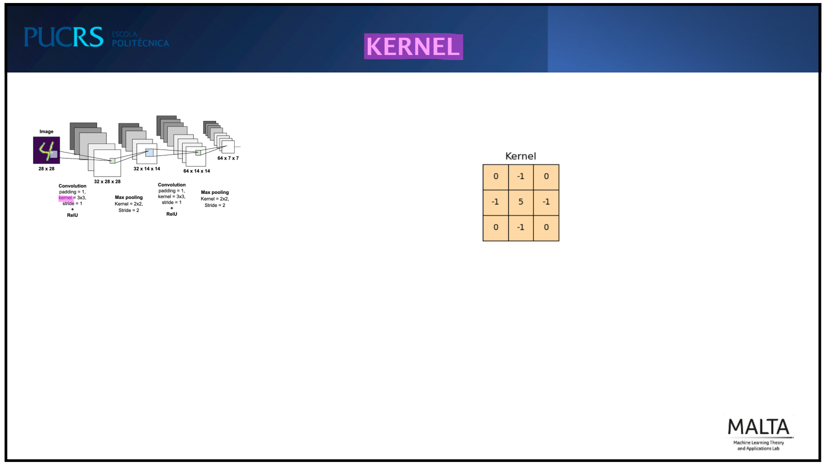
141



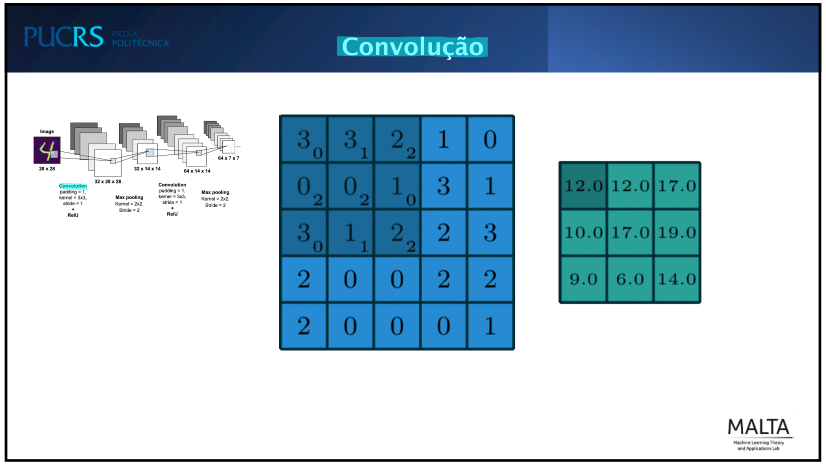
142



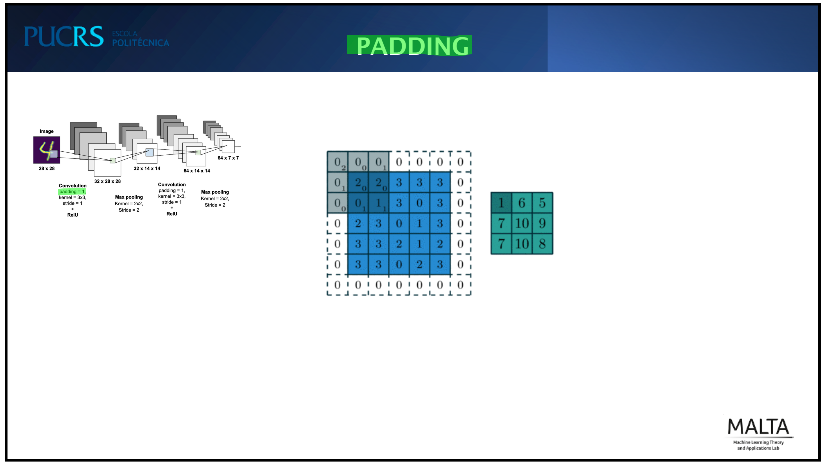
143



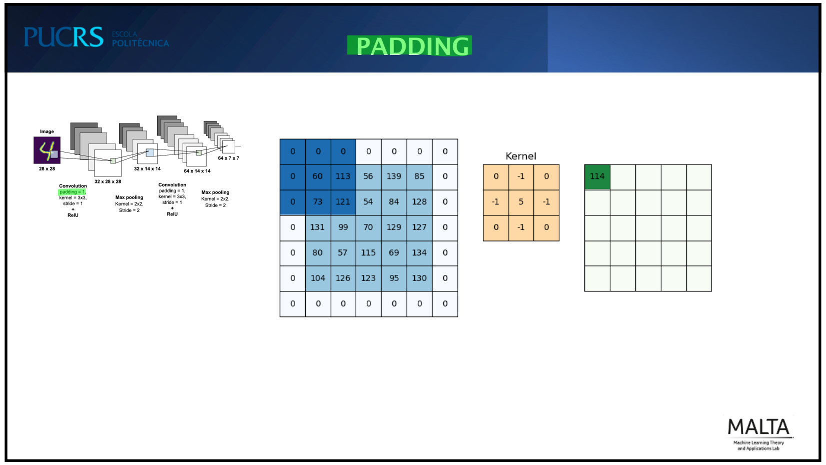
144



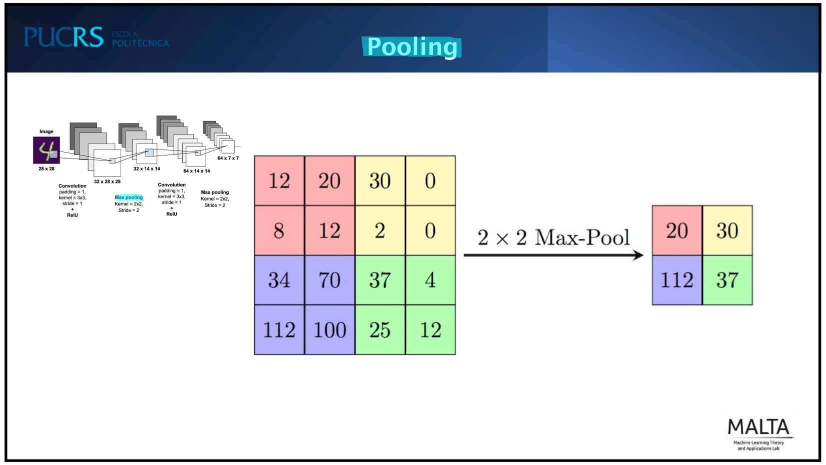
145



146



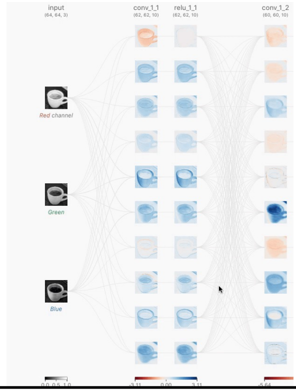
147



148

PUCRS ESCOLA POLITÉCNICA

USEM



Referência

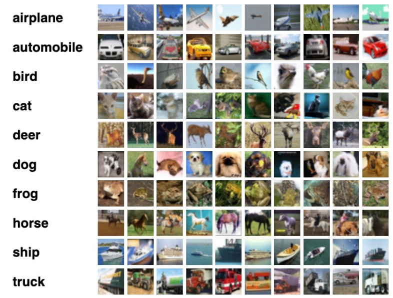
MALTA

149

PUCRS ESCOLA POLITÉCNICA

PRATICA

shorturl.at/ouL57

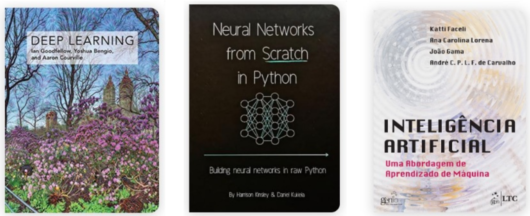


MALTA

150

PUCRS ESCOLA POLITÉCNICA

Referências



MALTA

151