MBA USP ESALQ

Basic concepts for Introduction to Deep Learning

Prof. Dr. Jeronymo Marcondes

Introduction

- Plan of attack:

1) Explain the concept
2) How does the data flow?
3) How is the optimization process?
4) Cost function and gradient descent
5) Practical part

Introduction

- Problem of classification and regression

$$y = f(X1, X2, \dots X3)$$

- Machine Learning is a technology in which computers has the ability to learn according to the expected answers through associations of different data, which can be images, numbers, and everything that this technology can identify.

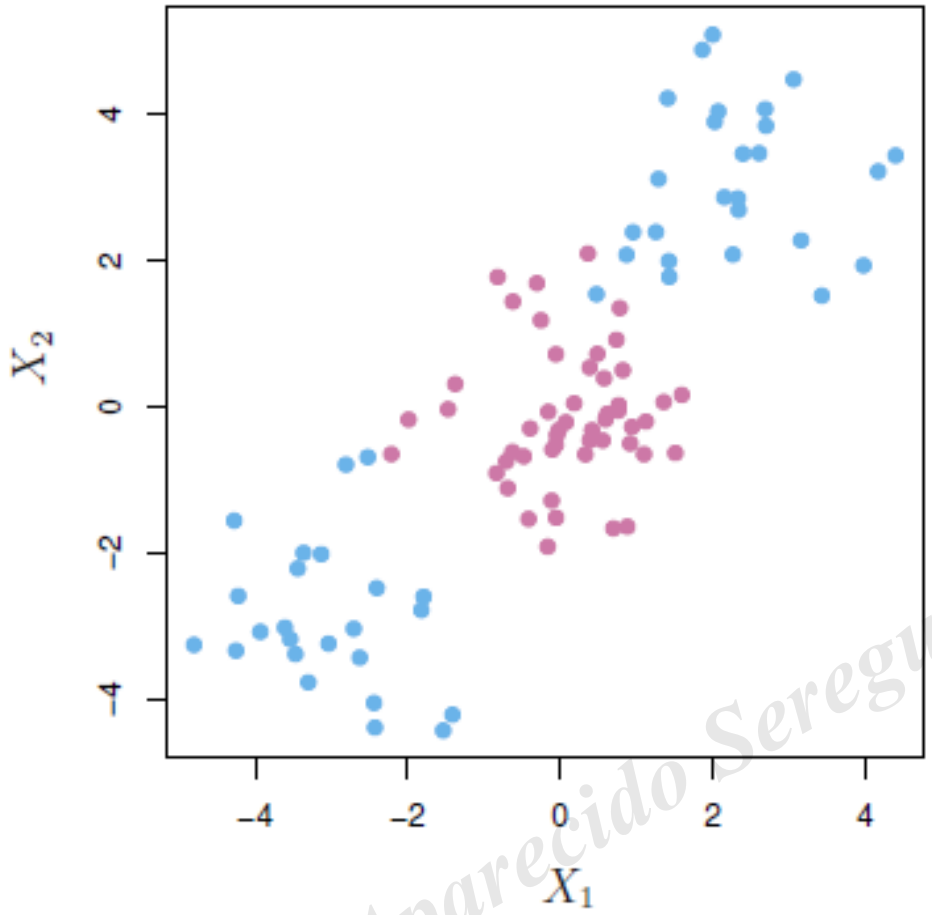
Examples

It is possible to perform models to:

classify customers into chance of not paying;

predict the value of a stock;

reduce a data set.



Morettin and Singer – Introduction to Data Science

Supervised AE

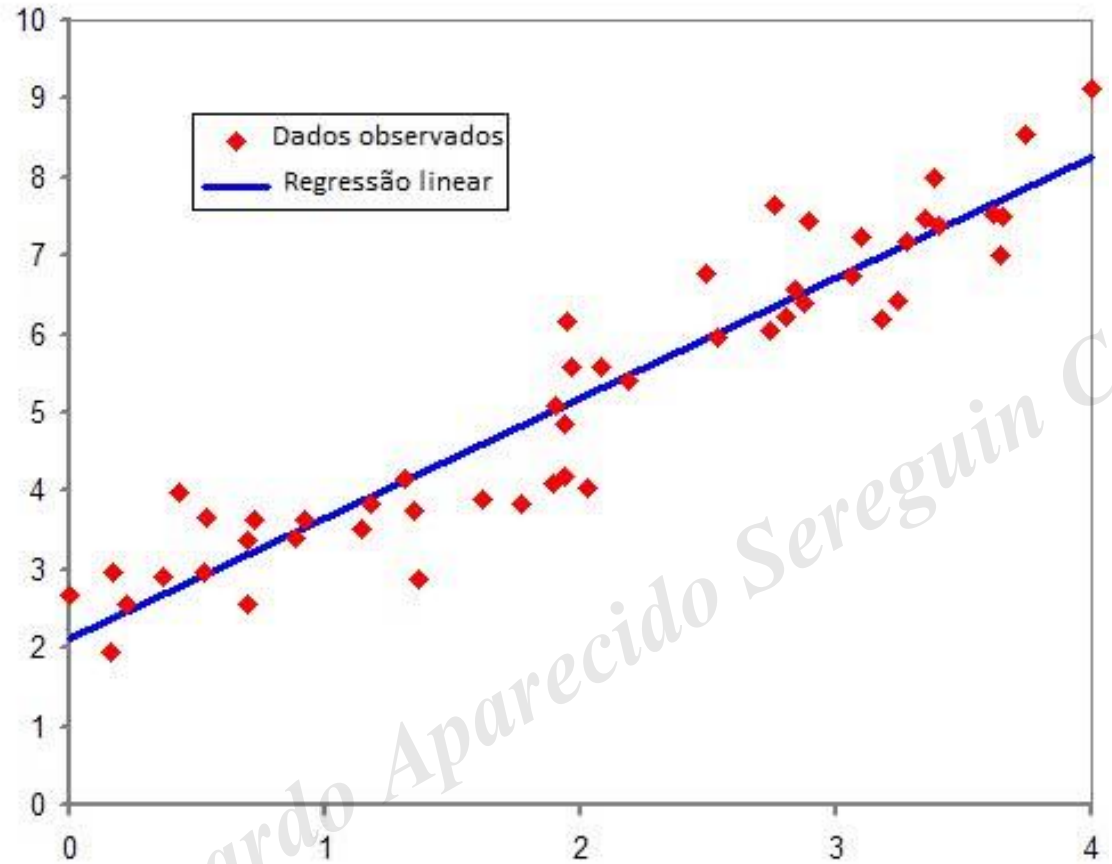
- Supervised problem
- Regression:

$$y = f(X) + e$$

$$y = a + bX + e$$

$$y = 0{,}2 + 0{,}1X$$

Supervised problem



Non-supervised AE

- There is only a set of predictive variables (inputs) and the objective is to describe associations and standards between these variables. In this case, there is not a response variable.

- Cluster Analysis and Principal Component Analysis.

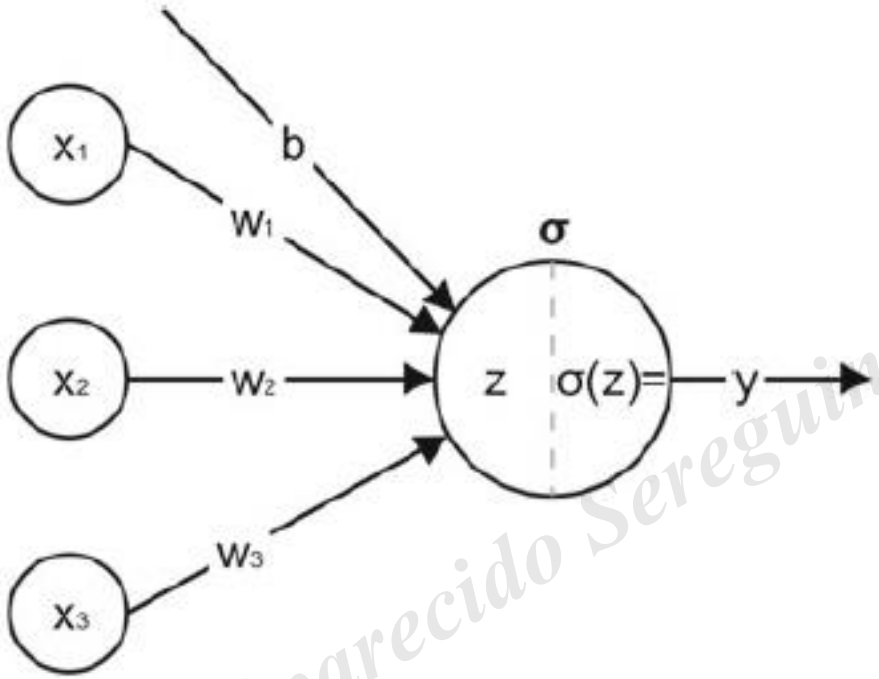
classification

- Response of the model is a qualitative variable.

- Example of default risk.

$$Default = f(Risk\ of\ the\ Individual)$$

Regression as ANN



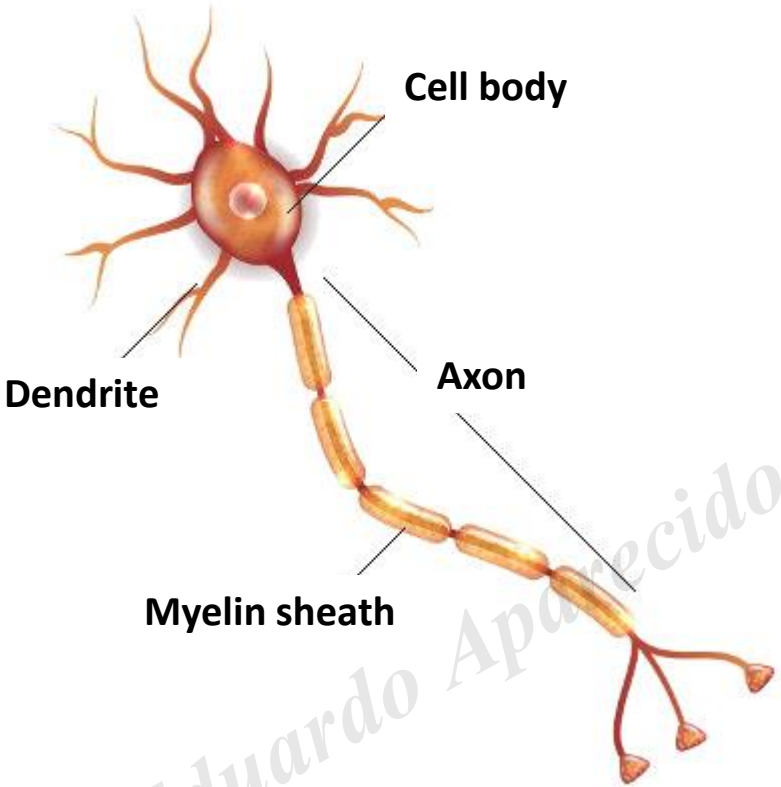
Artificial Neural Networks

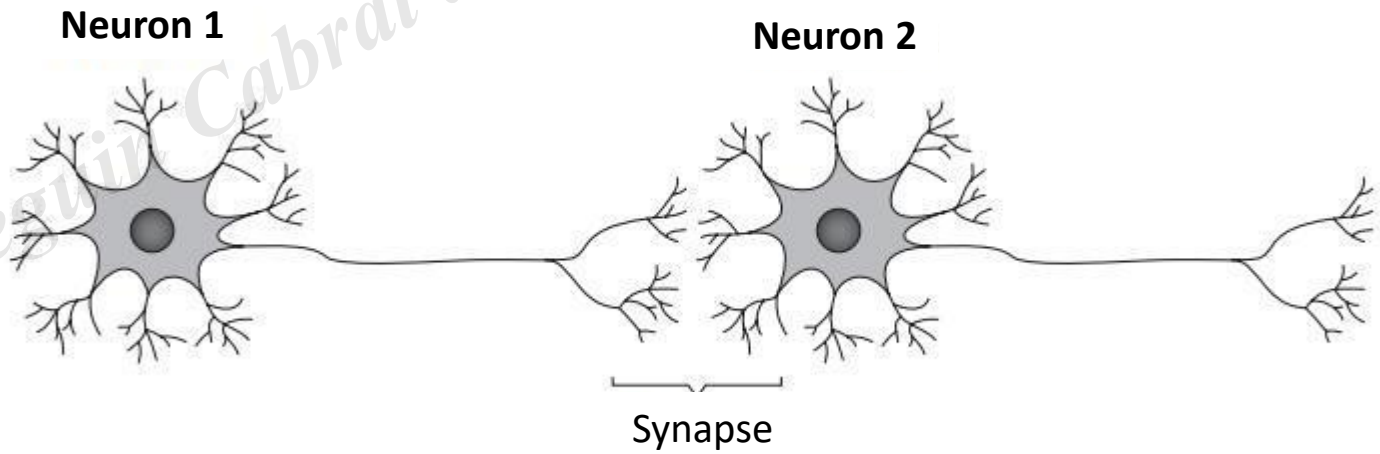
- Deep learning – concept

- In this class, we will start with the "shallow" neural network.

- The pioneer contributions to the Neural Networks (NN) area were from McCulloch and Pitts (1943), who introduced the idea of NN as computational machine, and from Hebb (1949), who postulated the first rule for organizational learning.

Artificial Neural Networks



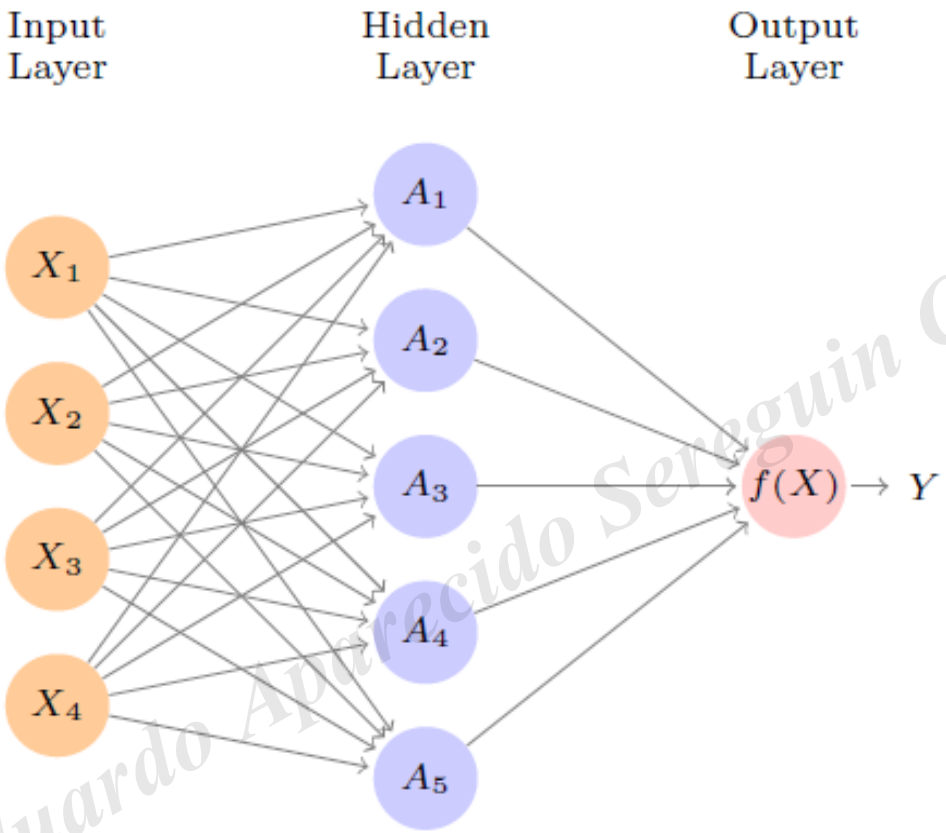


Artificial Neural Networks

- The idea of Rosenblatt's perceptron.
- Learn binary answers
- Trained weight to produce target vector

Shiffman (2012) demonstrates Perceptron as a computational model of a simple neuron, which consists in having several data inputs, a processing core and an output, which only allows the output of logical values.

Artificial Neural Networks



Artificial Neural Networks

- Neural network of the feedforward type.
- Neurons do not interconnect.
- Input layers cannot be smaller than the number of variables that we are using to explain Y.

How does it work?

- Each Neuron receives data from the inputs or from the previous lay.
- Differently from the regression case – it had only one input and output lay.
- Each Neuron when sending information to the next one multiplies the information by **weights.**

Mathematics

- When presented the value of n inputs:

$$X_1 = [X_{11}, X_{12} \dots X_{1n}]$$

These are our explanatory variables! For example, the customer's default risk:

$$X = [Did\ they\ paid?, Income, etc]$$

Mathematics

- When presented the value of output for t customers:

$$Y = [Y_1, Y_2, \ldots, Y_t]$$

These are our explained variables. In other words, if a certain customer has effectively paid.

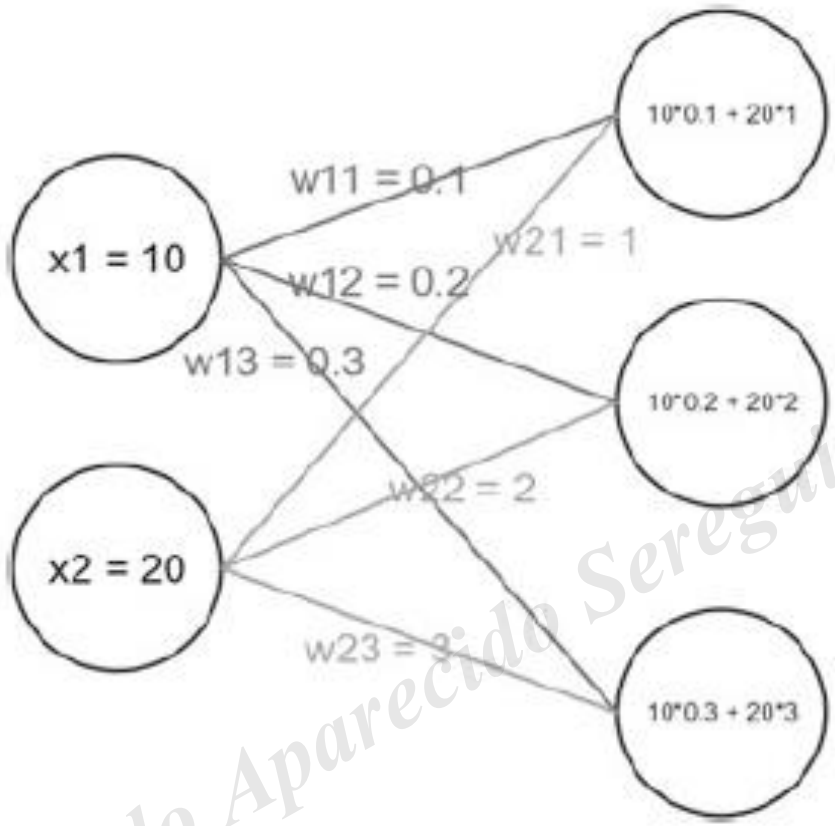
Mathematics

- When presented the value of weights for h neurons of the next layer:

$$w = [w_{11}, w_{12}, \dots, w_{nh}]$$

These are the values that will be multiplied by each input variable until the next hidden lay.

Visualization



Mathematics

- For each one of the neurons:

$$z = b + \sum_i w_i X_i$$

What is equivalent, in the case of the first neuron:

$$z_1 = b + w_{11} X_1 + w_{21} X_2$$

Activation function

- Suppose that we are with a problem of default classification with possible results 1 or 0.
- 1 will assign default and 0 won't.
- Our explained variable is dichotomic and it is a classification problem.

Activation function

Activation function processes the sign generated by the inputs and weights linear combination of synapses in order to generate the output sign of the neuron.

That is, it performs the “processing” of the information.

Linear and non-linear responses can appear.

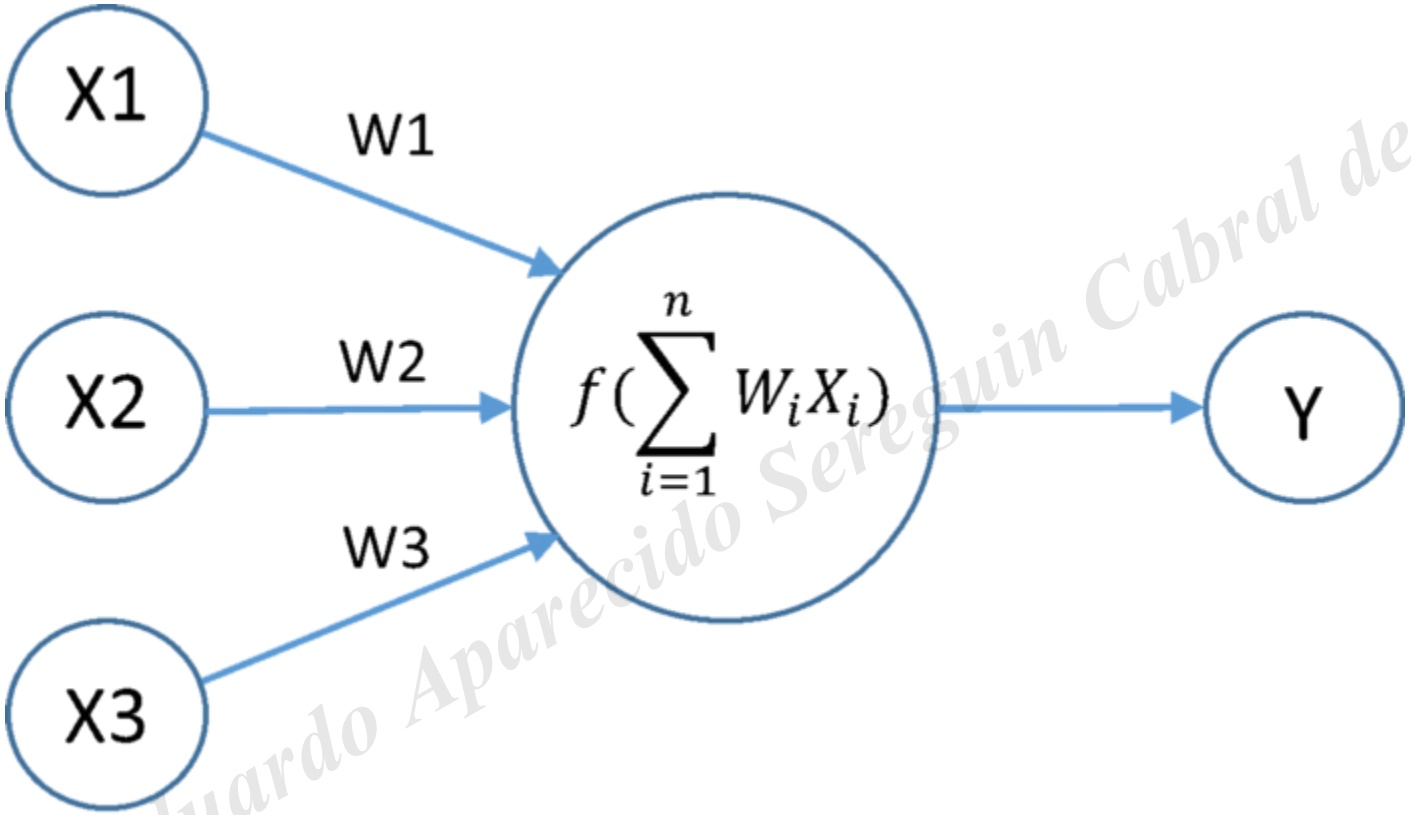
Activation function

- In order to decide the result from the previous neuron, we use (ReLU):

$$f(z) = \begin{Bmatrix} 1\ if\ z \geq 0 \\ 0\ otherwise \end{Bmatrix}$$

What is the difference compared to a regression?

Activation function



Activation function

$$z_1 = w_{11}X_1 + w_{21}X_2 = 1.2$$

$$Relu(z_1) = 1$$

Comparing with the table:

$$y - Relu(z_1)$$

What to do if it goes wrong??

Activation function

- Another example:

- Sigmoid function:

$$g(z) = \frac{e^z}{e^z + 1}$$

And if it was a linear function,

How does it work?

- The weight that is multiplied to the information depends on the origin/destiny pair.
- Result is inserted within an activation function.
- Depending on the result, such as in neuron, the sign is thrown to the next neuron.

How does it work?

- How to determine the weights?
- It is randomly initiated or with some pre-specified rule.
- Speculation case of the purchase values.

Explanation

- You know the quantity that you bought.
- You do not know the price per kg.
- You know the total that you spent.

$$gasto = ppk_{grape} q_{grape} + ppk_{apple} q_{apple}$$

Explanation

- Suppose that you spent 10.
- Received 1 kg of each.
- You guessed that the grape ppk is R$ 4.00 and the apple ppk is R$ R$ 5.00. You guess before knowing the total value.
- You were wrong for 10-9 = R$1.00. Residual error!

Explanation

- Suppose that you spent 10.
- Received 1 kg of each.
- You guessed that the grape ppk is R$ 4.00 and the apple ppk is R$ R$ 5.00. You guess before knowing the total value.
- You were wrong for 10-9 = R$1.00. Residual error!

Weights adjustment

- $erro\ residual_j = valor\ real_j - valor\ estimado\ _j$

- $\Delta w_j = \eta \nabla E$

Cost function

- Function that measures how much we are hitting.

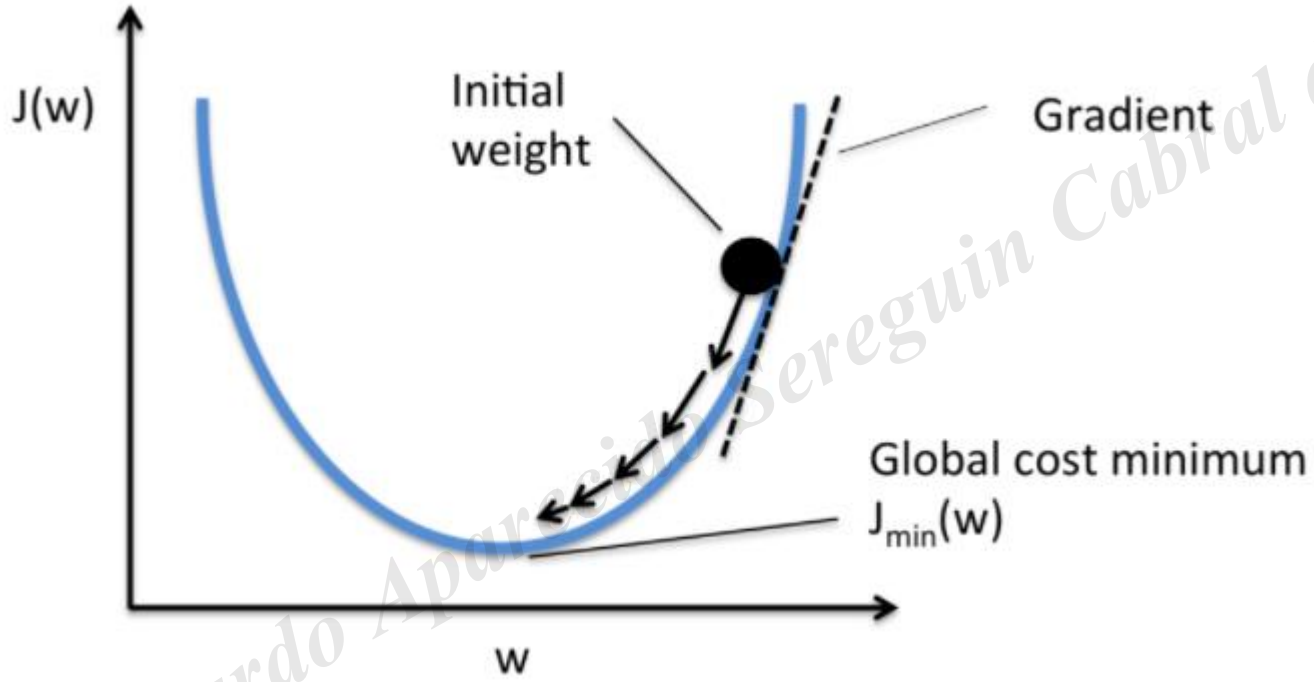
- For example: mean squared error.

$$E = \frac{1}{2} \sum_{\text{Trainning}} (t^n - y^n)^2$$

Cost function

	Estimated value	Real value	Error	Square of the error
Price 1	2	2	0	0
Price 2	3	5	-2	4
Price 3	2	4	-2	4
Price 4	5	1	4	16
Sum				24
				6

How is the adjust make?



Source: Deep Learning book

How is the adjust make?

- Guess an initial value for the weight
- Calculate the output values for these weights
- Use the adjustment formula to obtain new weights:

$$new\ weight = previous\ weight - \eta \nabla E$$

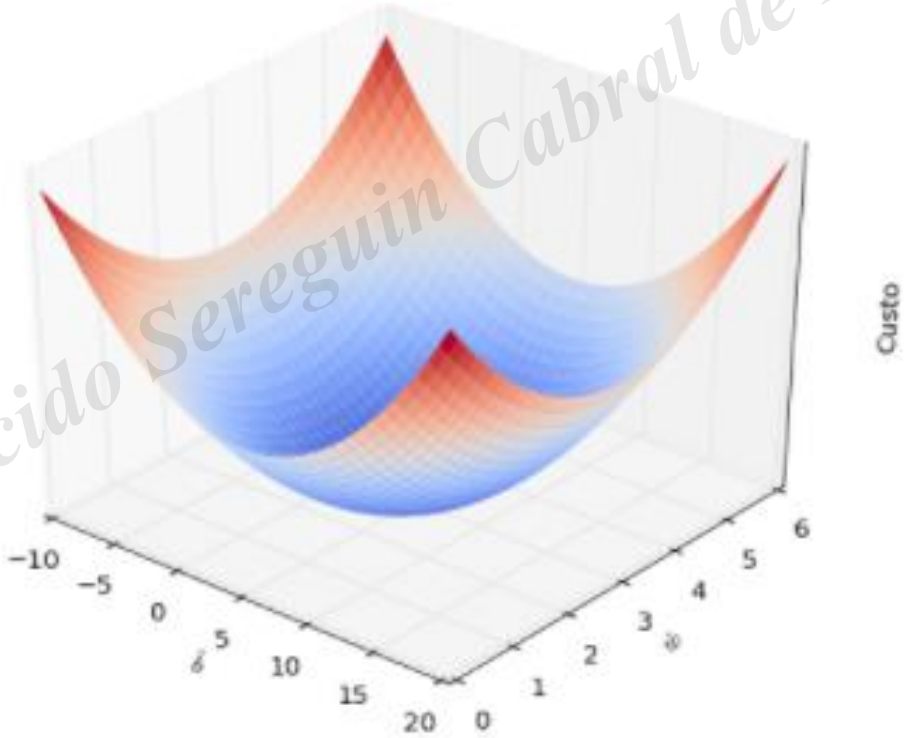
Gradient descent

- Process used to find the minimum of the cost function based on variation in weights.

- Through the backpropagation process, the errors obtained after the weight calculations got feedback in the network until we optimize the result.

Gradient descent

- Mathematically: what is the variation in the weights that most reduces my difference between the estimated value and the real value?



Source: https://matheusfacure.github.io/2017/02/20/MQO-Gradiente-Descendente/

Gradient descent

- Use of the differential calculation:

The gradient will provide variations in the loss function for small changes in the parameters:

$$\nabla(L) = \left[\frac{\partial L}{\partial w}, \frac{\partial L}{\partial b}\right]$$

Practical part