

## MÓDULO 2

# Aprendizagem supervisionada

Especialização em Machine Learning

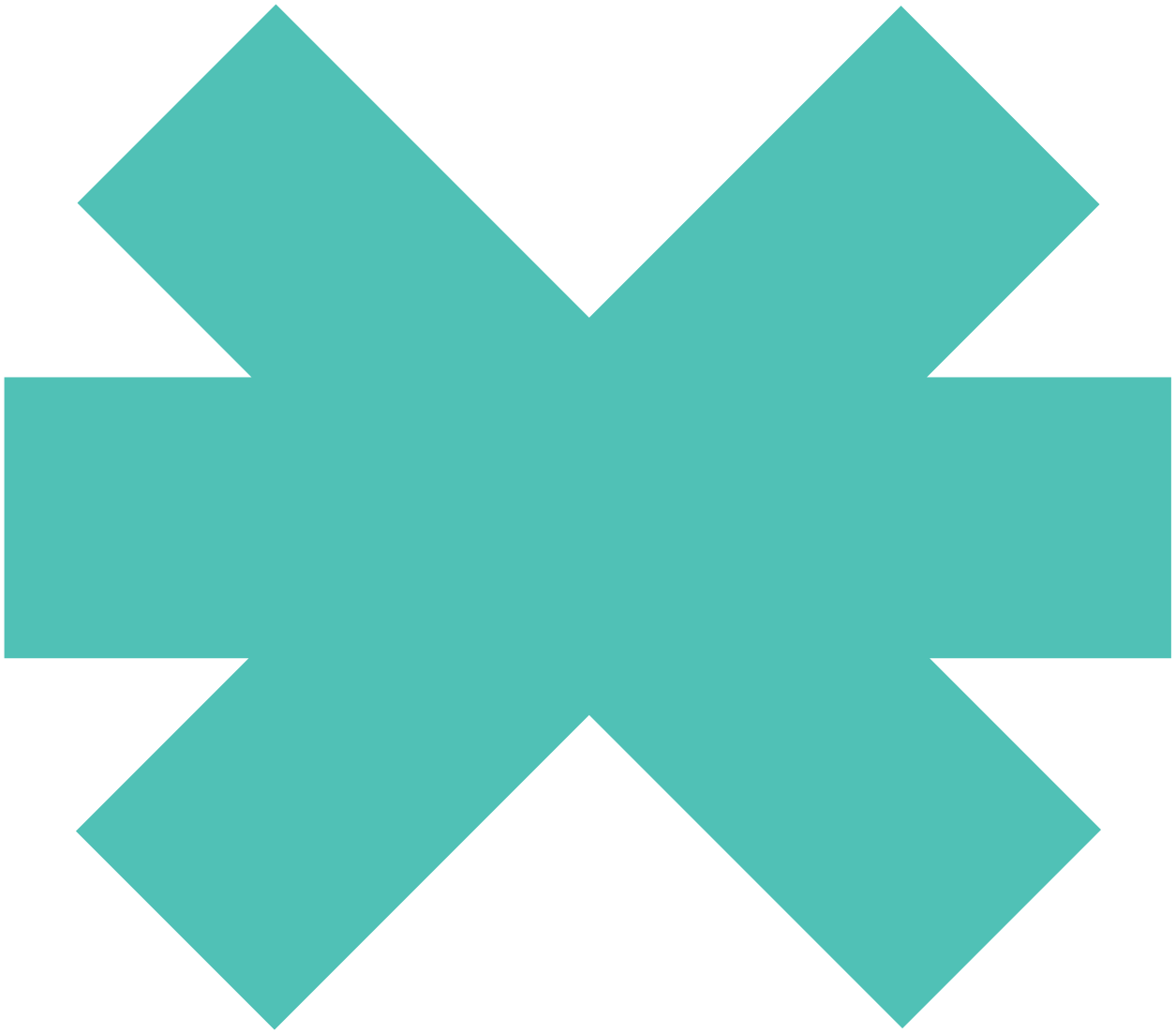
# 3

# Normalização



New  
Technology  
School

Tokio.



# 3 Normalização

## Sumário

3.1	Problema	04
3.2	Normalização	05
3.3	Algoritmo de treino atualizado	07

## 3.1 Problema

Habitualmente, as diferentes características num conjunto de dados apresentam distintos intervalos de valores, ou seja, algumas contam com valores em intervalos de unidades, outras de dezenas, centenas, milhares, etc.

Se usarmos *gradient descent*, estamos a calcular a inclinação em cada direção, coordenada ou para cada característica, e a avançar nessa direção.

Assim, os distintos intervalos afetam o treino, pois a amplitude de cada passo, controlada pelo rácio de aprendizagem e em função da amplitude da derivada parcial, para cada característica, pode ser muito diferente em cada direção.

Isto pode levar a que algumas características de intervalo superior “ocultem” outras de intervalo menor, ainda sendo estas últimas muito mais descritivas para modelar o conjunto de dados.

Exemplo:

$m \times n$	$x_{...0}$	$x_{...1}$	$x_{...2}$	$x_{...3}$	$x_{...4}$	$x_{...5}$	$x_{...6}$	$x_{...7}$	$x_{...8}$
$x_{0...}$	1	150	12	3.252	0	3.688	3587	0.012	0
$x_{1...}$	1	200	23	4.821	1	6.587	6215	0.045	0
$x_{2...}$	1	175	71	4.935	0	7.648	2178	0.313	0
$x_{3...}$	1	110	11	2.013	1	1.3459	9822	1.274	1

## 3.2 Normalização

O objetivo da normalização, como o seu próprio nome indica, é “normalizar” o conjunto de dados, transformando as características noutras, que apresentem um intervalo comum.

Logo, pré-processaremos as características antes de as usar no modelo, com o objetivo de conseguir que a aprendizagem seja mais equilibrada em todas as dimensões e tenha em conta todas as características, em função da sua importância, para determinar a variável dependente/objetivo  $Y$ .

Existem vários métodos de normalização que poderemos selecionar, segundo as circunstâncias do modelo e o conjunto de dados. De um modo geral, pretendemos que os valores permaneçam num intervalo único  $[-1, 1]$  e com média aprox. de 0.

Para isso, podemos usar métodos como normalizar segundo a variância, o intervalo mínimo e máximo dos valores, etc.

Neste caso, normalizamos segundo a média e variância originais do conjunto de dados:

$$x = \frac{x - \mu_{x_j}}{\sigma_{x_j}}$$

Figura 11

Código Latex:

```
x = \frac{x - \mu_{x_j}}{\sigma_{x_j}}
```

Recordemos que não devemos normalizar  $x_y$  uma vez que o seu valor deve ser sempre 1, para vetorizar a operação.

Exemplos:

$Y$	$x_{...1}$	$x_{...2}$	$x_{...3}$	$x_{...4}$
100000	100	3	1	1000
120000	110	4	2	1200
50000	80	2	1	400
180000	120	4	2	5600
175000	120	4	2	2400
200000	160	3	2	2400
100000	90	2	1	30
75000	80	2	1	0
125000	110	3	2	3600
120000	100	2	1	3600

	107,0000	2,9000	1,5000	2023,0000
	23,5938	0,8756	0,5270	1840,1573
$Y$	$x_{\rightarrow 1}$	$x_{\rightarrow 2}$	$x_{\rightarrow 3}$	$x_{\rightarrow 4}$
100000	-0,2967	0,1142	-0,9487	-0,5559
120000	0,1272	1,2563	0,9487	-0,4472
50000	-1,1444	-1,0279	-0,9487	-0,8820
180000	0,5510	1,2563	0,9487	1,9439
175000	0,5510	1,2563	0,9487	0,2049
200000	2,2464	0,1142	0,9487	0,2049
100000	-0,7205	-1,0279	-0,9487	-1,0831
75000	-1,1444	-1,0279	-0,9487	-1,0994
125000	0,1272	0,1142	0,9487	0,8570
120000	-0,2967	-1,0279	-0,9487	0,8570

## 3.3 Algoritmo de treino atualizado

O algoritmo de treino completo de um problema, de regressão linear múltipla, atualizado com normalização, seria:

1. Compilar os exemplos  $X$  e os seus resultados previamente conhecidos  $Y$ .
2. Normalizar os exemplos  $X$ .
3. Escolher um rácio de aprendizagem  $\alpha$ .
4. Inicializar os pesos  $\Theta$ , de forma aleatória.
5. Iterativamente, calcular o custo e suas derivadas/ inclinações e atualizar  $\Theta$ .
6. Finalizar quando  $\Theta$  converja num valor ótimo.
7. Modificar o rácio de aprendizagem  $\alpha$  se necessário.
8. Obter a  $\Theta$  ótima e realizar predições com ela.

