

# **Introdução a Regressão Logística**

# Atividade de Leitura

Seções 4-4.3 do  
**Introduction to Statistical  
Learning**  
por Gareth James, et al.

# Background

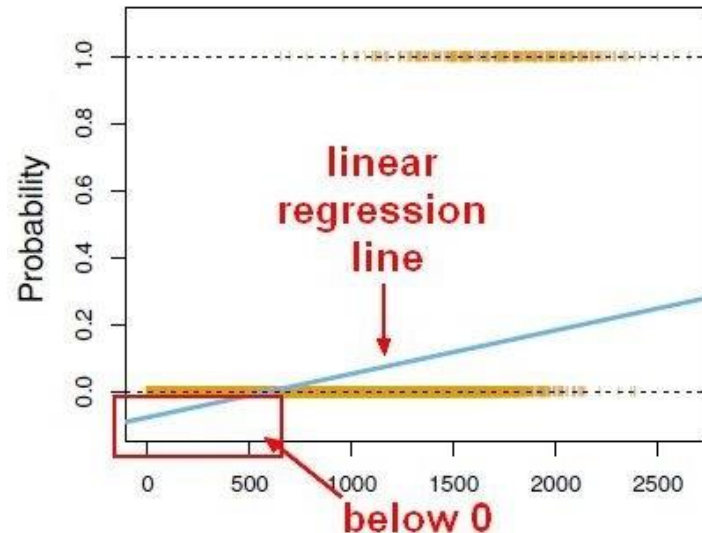
- Queremos estudar sobre Regressão Logística como um método para **Classificação**.
- Alguns exemplos de problemas de classificação:
  - Spam versus e-mails Legítimos
  - Inadimplência (sim/não)
  - Diagnóstico de Doenças
- Todos os exemplos acima são de Classificação Binária.

# Background

- Até agora, vimos apenas problemas de regressão em que tentamos prever um valor contínuo.
- Embora o nome possa ser confuso no início, a regressão logística nos permite resolver problemas de classificação, onde estamos tentando prever categorias discretas.
- A convenção para classificação binária é ter duas classes: 0 e 1.

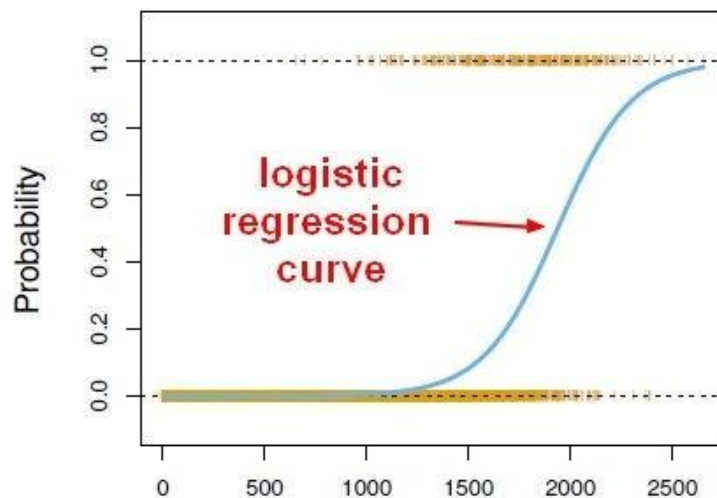
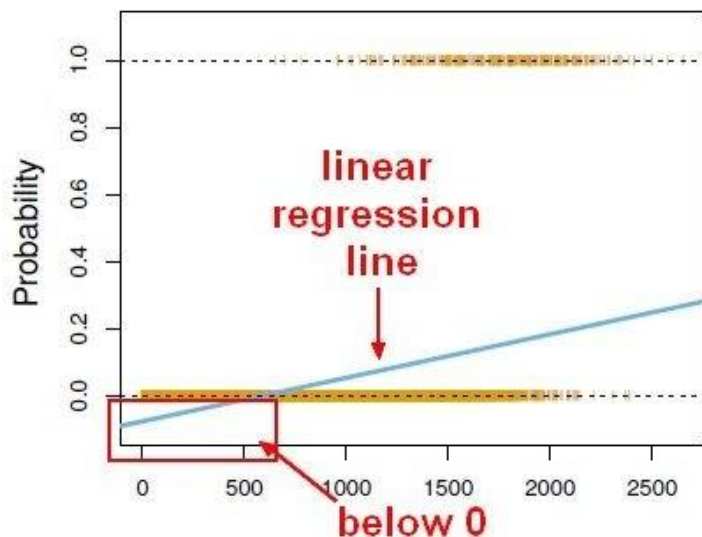
# Background

- Não podemos usar um modelo de regressão linear normal em grupos binários. Não obteremos um bom ajuste:



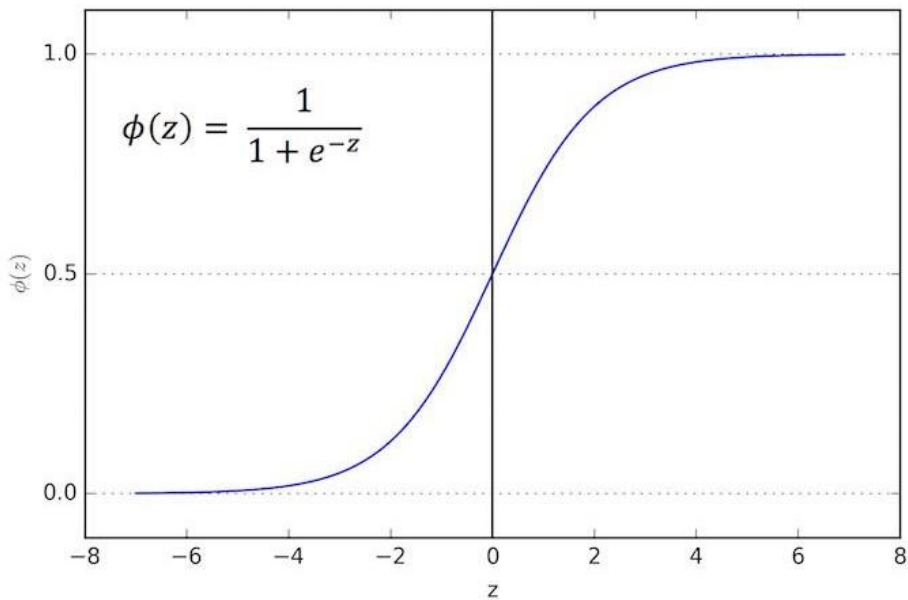
# Background

- Em vez disso, podemos transformar nossa regressão linear em uma curva de regressão logística.



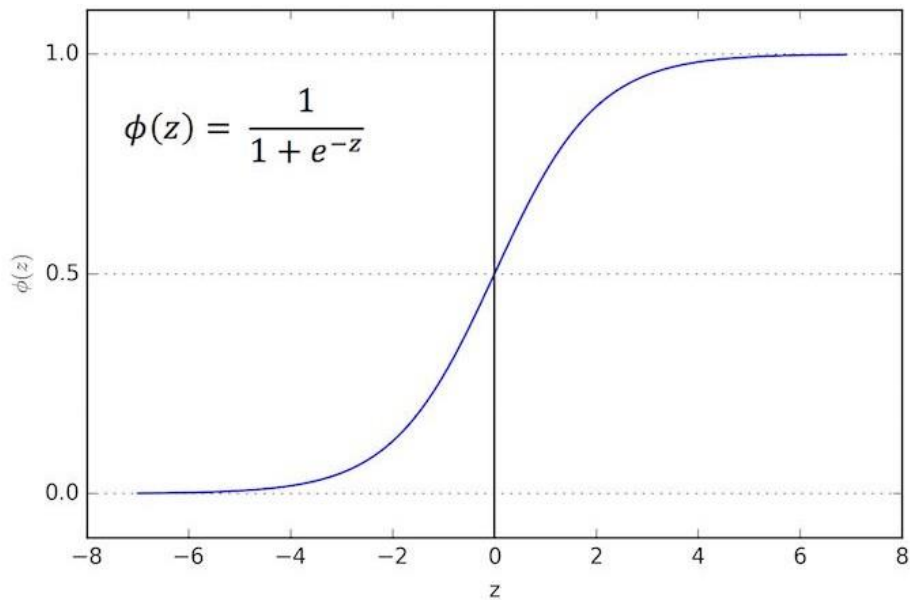
# Sigmoid Function

- A função sigmóide (também conhecida como logística) recebe qualquer valor e retorna outro valor entre 0 e 1.



# Sigmoid Function

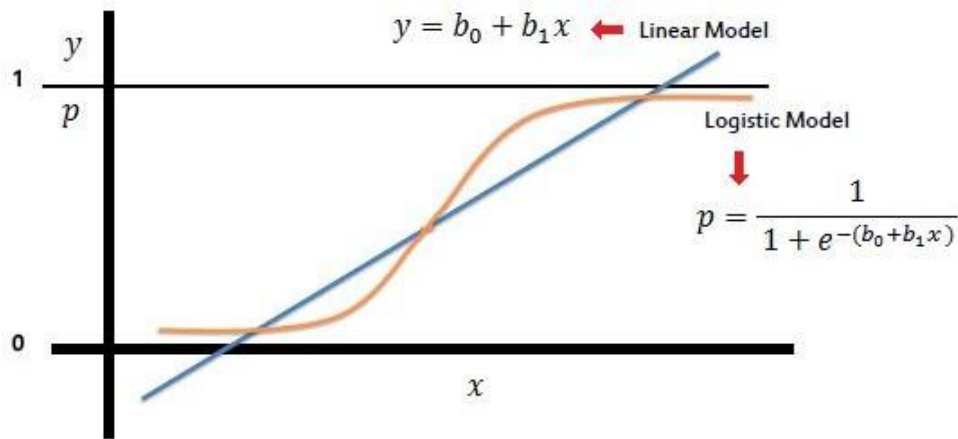
- Isso significa que podemos pegar nossa solução de regressão linear e colocá-la na função sigmóide.





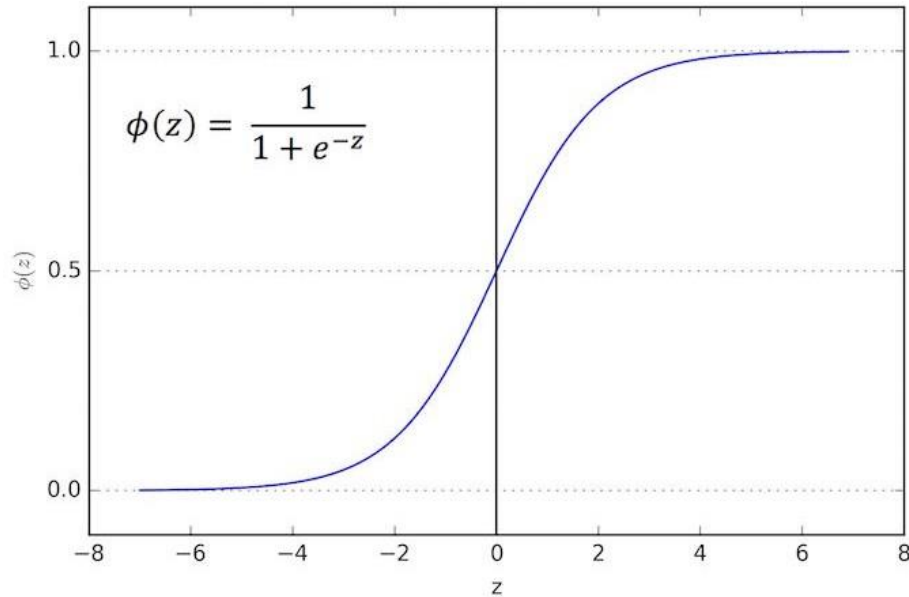
# Sigmoid Function

- Isso significa que podemos pegar nossa solução de regressão linear e colocá-la na função sigmóide.



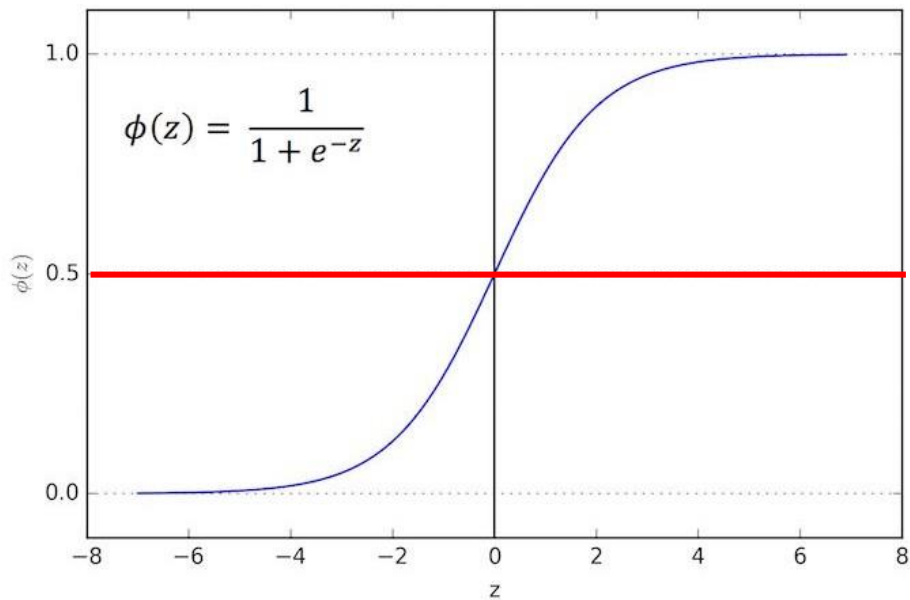
# Sigmoid Function

- Isso resulta em uma probabilidade entre 0 e 1 de pertencer à classe 1.



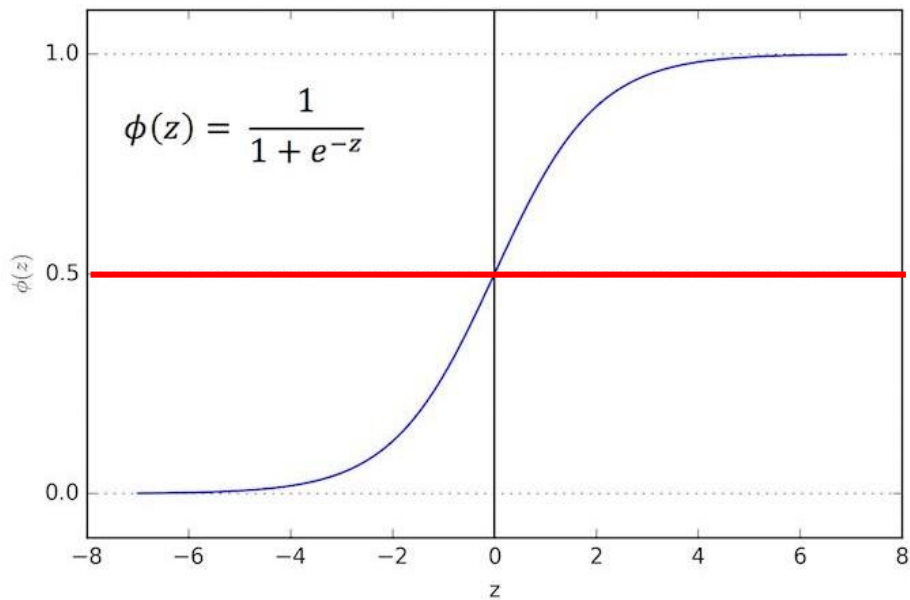
# Sigmoid Function

- Podemos definir um ponto de corte em 0.5, qualquer coisa abaixo resulta em classe 0, qualquer coisa acima é classe 1.



# Análise

- Usamos a função logística para produzir um valor entre 0 e 1. Com base nessa probabilidade, atribuímos o resultado a uma classe.



# Avaliação do Modelo

- Depois de treinar um modelo de regressão logística em alguns dados de treinamento, você avaliará o desempenho do seu modelo em alguns dados de teste.
- É possível utilizar uma matriz de confusão para avaliar modelos de classificação.

# Avaliação do Modelo

- Por exemplo, considere o caso da realização de testes para determinada doença.

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Exemplo: Teste para presença de doença  
NO = teste negativo = False = 0  
YES = teste positivo = True = 1

# Matriz de Confusão

n=165		Predicted: NO	Predicted: YES	
Actual: NO		TN = 50	FP = 10	60
Actual: YES		FN = 5	TP = 100	105
		55	110	

Terminologia Básica:

- **Verdadeiros Positivos**  
(True Positives - TP)
- **Verdadeiros Negativos**  
(True Negatives - TN)
- **Falsos Positivos**  
(False Positives - FP)
- **Falsos Negativos**  
(False Negatives - FN)

# Matriz de Confusão

n=165	Predicted: NO	Predicted: YES	
	Actual: NO	TN = 50	FP = 10
Actual: YES	FN = 5	TP = 100	105
	55	110	

Acurácia:

- No geral, com que frequência o modelo está **correto**?
- $(TP + TN) / \text{total} = 150/165 = 0.91$



# Matriz de Confusão

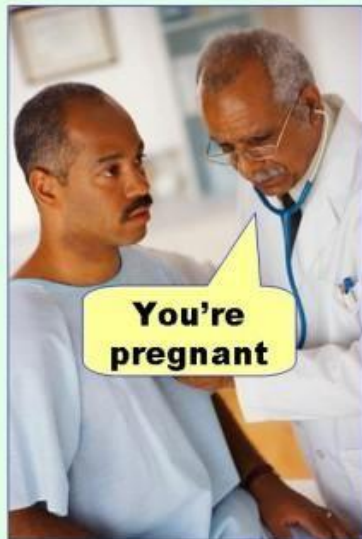
n=165		Predicted: NO	Predicted: YES	
Actual: NO		TN = 50	FP = 10	60
Actual: YES		FN = 5	TP = 100	105
		55	110	

Taxa de Classificação  
Incorreta (Taxa de Erro):

- No geral, com que frequência o modelo está **errado**?
- $(FP + FN) / \text{total} = 15/165 = 0.09$

# Matriz de Confusão

**Type I error**  
(false positive)



**Type II error**  
(false negative)

