



INSTITUTO DO EMPREGO E FORMAÇÃO PROFISSIONAL, IP
DELEGAÇÃO REGIONAL DE LISBOA E VALE DO TEJO
CENTRO DE EMPREGO E FORMAÇÃO PROFISSIONAL DE SINTRA

UFCD – 10810

Fundamentos do desenvolvimento de modelos analíticos em Python

3 – Regressão Logística

Carga horária: 25 horas

Formador: Manuel Viana

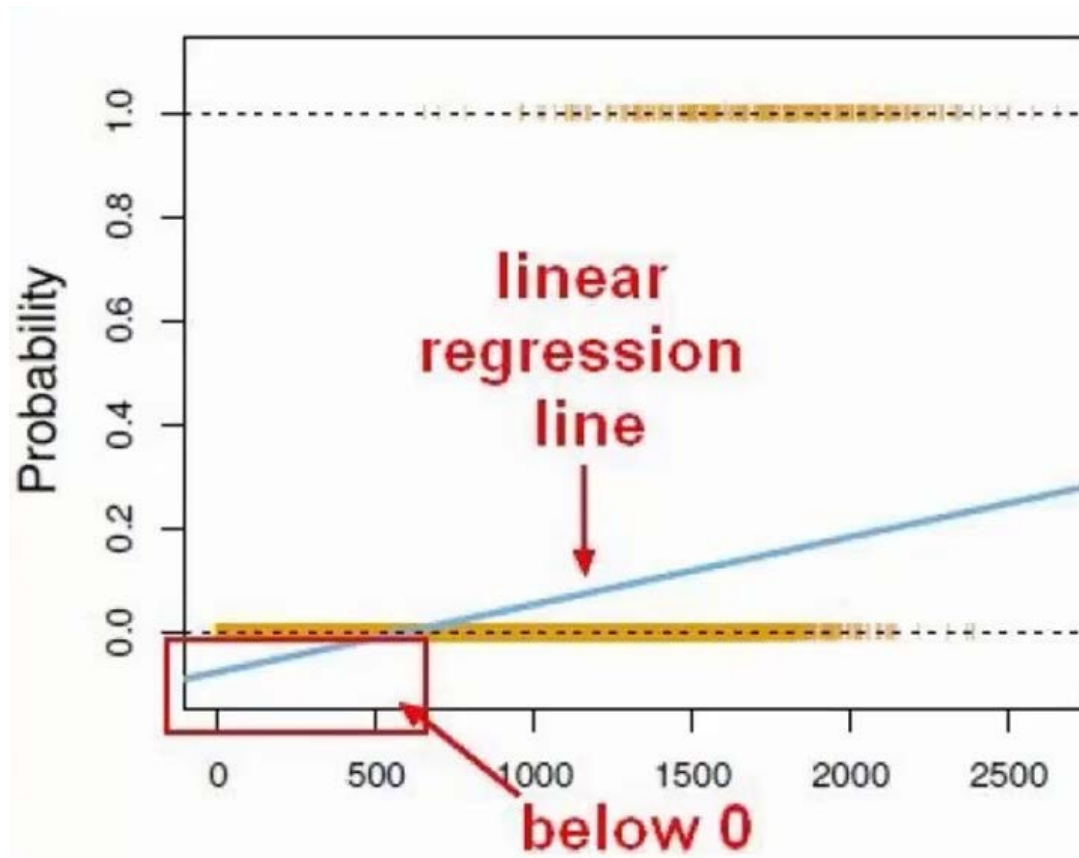


O que é Regressão Logística?

- A Regressão Logística é um método de classificação.
- Alguns exemplos de problemas de classificação:
 - Filtros de emails (spam ou não)
 - Modelos preditores de clientes inadimplentes (que não pagam as suas dívidas)
 - Diagnóstico de doenças
- Permite a resolução de problemas de classificação, onde se tentam prever categorias discretas.
- Por convenção, a classificação é binária – duas classes: 0 e 1 (Falso e Verdadeiro)

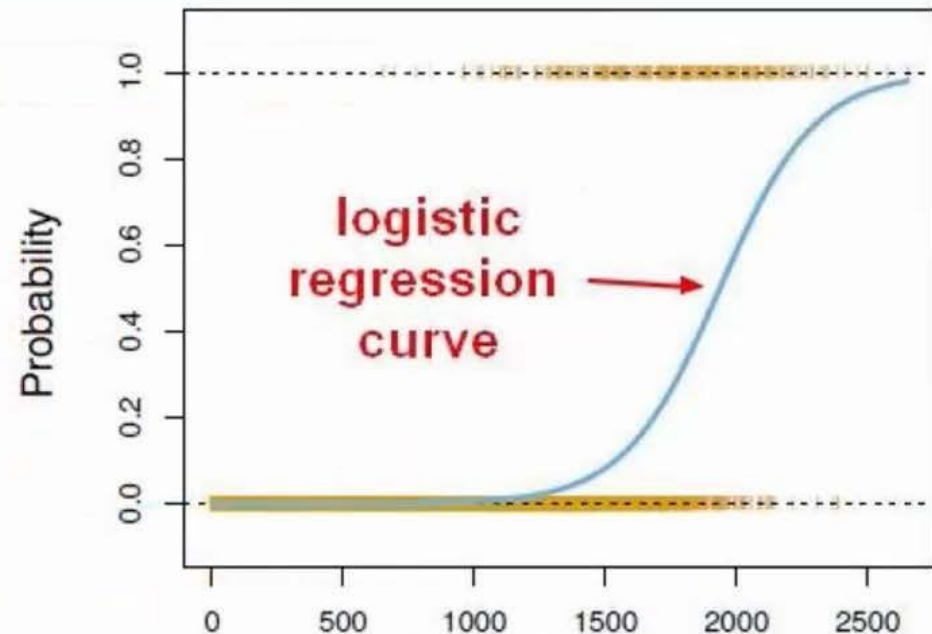
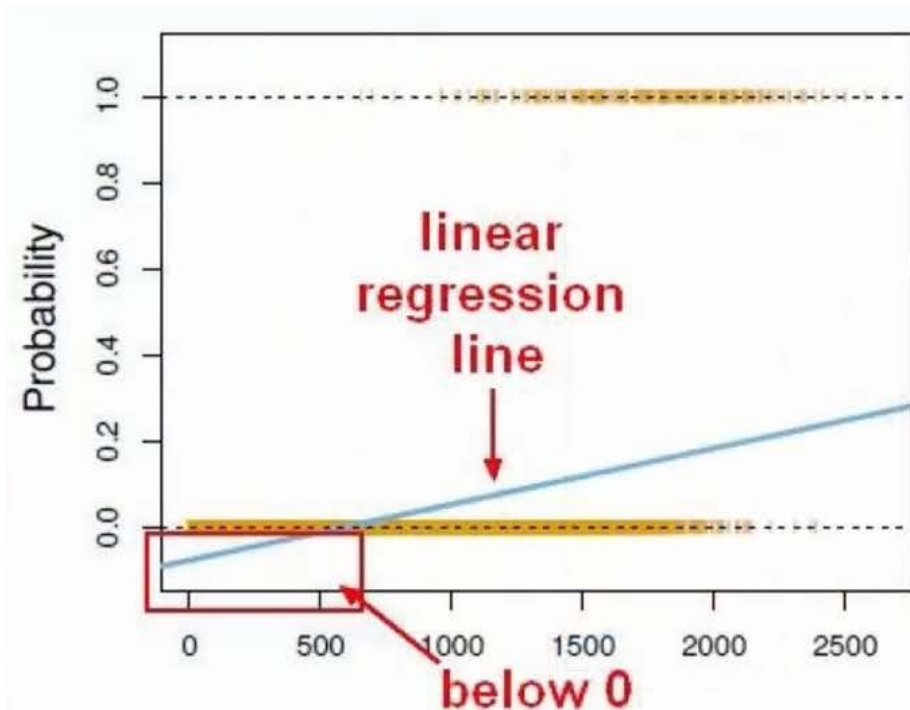
Porque é que não se utiliza a Regressão Linear para resolver problemas de classificação?

- A Regressão Linear permite probabilidades acima de 100% ou abaixo de 0%.
- Não consegue classificar conjuntos de dados.



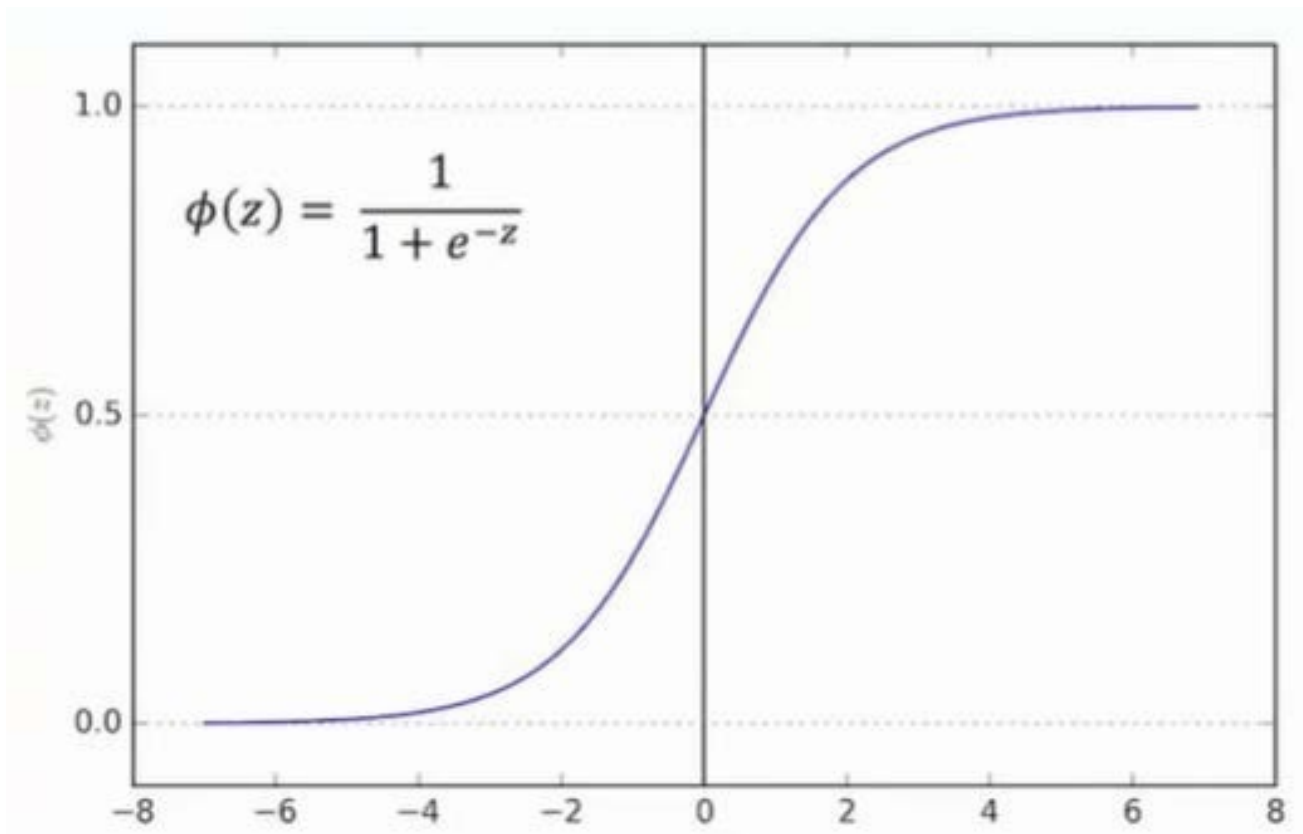
Regressão Logística

- A regressão logística contém uma função específica que retorna sempre um valor entre 0 e 1, independentemente dos valores atribuídos à função.



Regressão Logística

- A função logística (também conhecida por sigmóide) só retorna valores entre 0 e 1.
 - Estes são interpretados com a probabilidade de pertencerem à classe 0 ou 1.



Regressão Logística

- Após treinarmos um modelo de regressão logística, testa-se o mesmo num conjunto de dados de teste. Contudo, as métricas de avaliação de desempenho do modelo são diferentes das métricas utilizadas na Regressão Linear.
- A forma de avaliar o modelo de regressão logística, passa pela utilização de uma **matriz**.
- **Exemplo:** Teste para a presença de uma doença
 - NÃO: falso – 0
 - SIM: Verdadeiro – 1

| | Valor predito: NÃO | Valor predito: SIM |
|--------------------|-----------------------|-----------------------|
| Valor correto: NÃO | 50 | 10 |
| Valor correto: SIM | 5 | 100 |

Regressão Logística

- **Exemplo:** Teste para a presença de uma doença
 - NÃO: falso – 0
 - SIM: Verdadeiro – 1

| | Valor predito: NÃO | Valor predito: SIM |
|--------------------|-----------------------|-----------------------|
| Valor correto: NÃO | 50 (TN) | 10 (FP) |
| Valor correto: SIM | 5 (FN) | 100 (TP) |

- Terminologia:
 - True Positive (TP)
 - True Negative (TN)
 - False Positive (FP)
 - False Negative (FN)

Regressão Logística

- **Exemplo:** Teste para a presença de uma doença

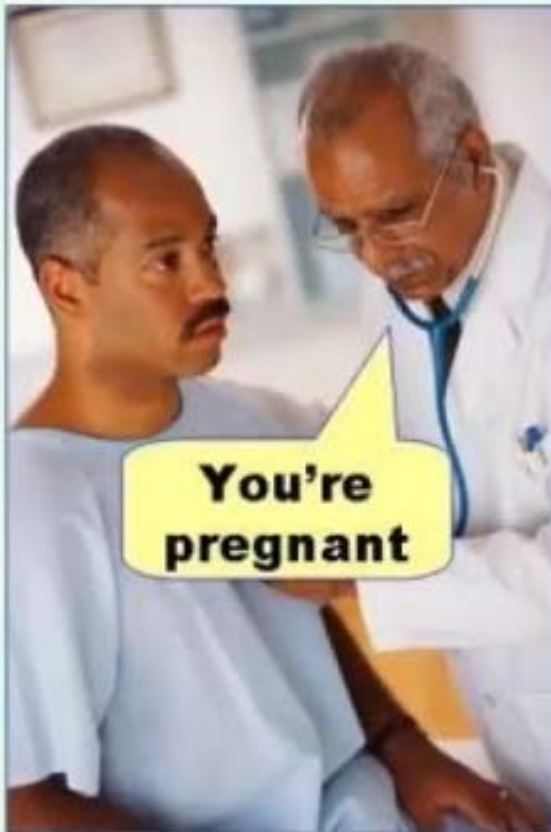
- NÃO: falso – 0
- SIM: Verdadeiro – 1

| | Valor predito: NÃO | Valor predito: SIM |
|--------------------|-----------------------|-----------------------|
| Valor correto: NÃO | 50 (TN) | 10 (FP) |
| Valor correto: SIM | 5 (FN) | 100 (TP) |

- Precisão:
- Classificações corretas - Qual a taxa de acertos do modelo?
 $(TP + TN) / \text{total} = 150 / 165 = \mathbf{0.91 (91\%)}$
- Classificações incorretas - Qual a taxa de erro do modelo?
 $(FP + FN) / \text{total} = 15 / 165 = \mathbf{0.09 (9\%)}$

Regressão Logística

Type I error
(false positive)



Type II error
(false negative)

