

Técnicas Estatísticas de Predição

Otaviano Francisco Neves



Regressão logística



Modelo de Regressão Logística

- O modelo de **regressão logística** é semelhante ao modelo de **regressão linear**.
- É um modelo de regressão múltipla em que a variável resposta (Y) é qualitativa e dicotômica: Exemplo: (sim/não) , (sucesso / fracasso), (0 ou 1);
- As variáveis explicativas podem ser quantitativas ou qualitativas.



Probabilidade

- O modelo de regressão logístico é utilizado quando a variável resposta é qualitativa com dois resultados possíveis. (Sucesso e Fracasso)
- Probabilidade de sucesso = p
- Probabilidade de fracasso = $1-p$



Chance

- $\text{Chance} = (\text{prob. de sucesso}) / (\text{prob. de fracasso})$
- Por exemplo, se a probabilidade de sucesso é 0,75, a chance é igual a:

$$p / (1-p) = 0,75 / 0,25 = 3$$



Razão de Chance

- Considere Grupo A
- $\text{Chance A} = (\text{probabilidade de sucesso em A}) / (\text{probabilidade de fracasso em A})$
- Considere Grupo B
- $\text{Chance B} = (\text{probabilidade de sucesso em B}) / (\text{probabilidade de fracasso em B})$
- $\text{Razão de Chance (A/B)} = \text{Chance A} / \text{Chance B}$



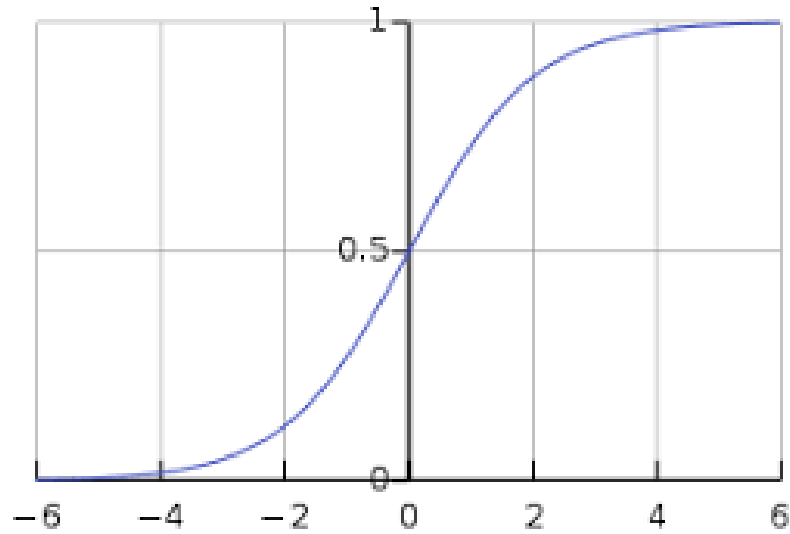
Cálculo

Grupo	Sucesso	Fracasso	Total
A	52	39	91
B	43	44	87
Total	95	83	178

- Chance de Sucesso no Grupo A: $p1 / (1-p1) = (52/91) / (39/91) = 0,57 / 0,43 = 1,33$
- Chance de Sucesso no Grupo B : $p2 / (1-p2) = (43/87) / (44/87) = 0,49 / 0,51 = 0,96$
- Razão de chances de Sucesso entre o grupo A, em relação ao Grupo B : $[p1 / (1- p1)] / [p2 / (1- p2)] = 1,33 / 0,96 = 1,39$



Modelo - logito



$$y = \frac{e^{ax+b}}{1 + e^{ax+b}}$$



Modelo – Coeficientes

Para uma regressão logística, a razão de chances para a observação i é dada por

$$r_i = \frac{p_i}{1 - p_i} = e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}}$$



Interpretação dos Coeficientes

- Considerando uma variável x_{1i} , e^{β_1} indica o aumento (ou redução) da razão de chances quando aumentamos em uma unidade a variável x_{1i} ;
- Se x_{1i} for uma variável “dummy” (categórica), o termo e^{β_1} indica o quanto a razão de chances se altera quando passa de uma categoria para outra.



Classificação com Regressão Logística

	Classificação 0 observada	Classificação 1 observada
Classificação 0 predita	Verdadeiro negativo	Falso negativo
Classificação 1 predita	Falso positivo	Verdadeiro positivo



Exemplo

Dados - Titanic





Exemplo – Dados Titanic IA



Titanic

- O naufrágio do **Titanic** é um dos acidentes mais trágicos da história. Em 15 de abril de 1912, em sua viagem inaugural, o **Titanic** afundou depois de colidir com um iceberg, matando 1502 de 2224 passageiros e tripulantes.



Dados

- Passageiro : Sequencial
- Sobrevivente : (Sobreviveu?): 0 = Não 1 = Sim
- Classe: Tipo da passagem 1 = 1ª Classe, 2 = 2ª Classe 3 = 3ª Classe
- Nome : Nome do passageiro
- Sexo: Sexo do passageiro
- Idade: Idade do passageiro
- Irmãos : Quantidade de irmãos / cônjuges a bordo do Titanic
- Pais : Quantidade de pais / crianças a bordo do Titanic
- Tarifa: Valor da passagem
- Informações retiradas do site do Kaggle (<https://www.kaggle.com/c/titanic/data>)



Cálculo da “Odds”

Grupo	1	0	Total
Feminino	197	64	261
Masculino	93	360	453
Total	290	424	714

- Chance de Sobrevivência no Grupo Feminino:

$$p1 / (1-p1) = (197/64) = 3,0783$$

- Chance de Sucesso no Grupo Masculino :

$$p2 / (1-p2) = (93/360) = 0,2583$$

- Razão de Chances (“odds”) de Sobrevivência entre o Grupo Feminino , em relação ao Grupo Masculino :

$$[p1 / (1- p1)] / [p2 / (1- p2)] = 3,0783 / 0,2583 = 11,92$$



Aplicação do modelo logístico



Questões



Classifique cada variável;



Faça uma análise descritiva;



Faça uma análise de regressão logística com as variáveis;

Interprete a tabela de Análise de Variância
Interprete os parâmetros
Interprete a qualidade do modelo



Calcule a predição para os indivíduos. determine o ponto de corte e calcule a taxa de acerto do modelo.