



PUC Minas

DIRETORIA DE
EDUCAÇÃO CONTINUADA

Pós Graduação *Lato Sensu*

**CIÊNCIA DE DADOS E
BIG DATA**

Técnicas Estatísticas de Predição

Programa

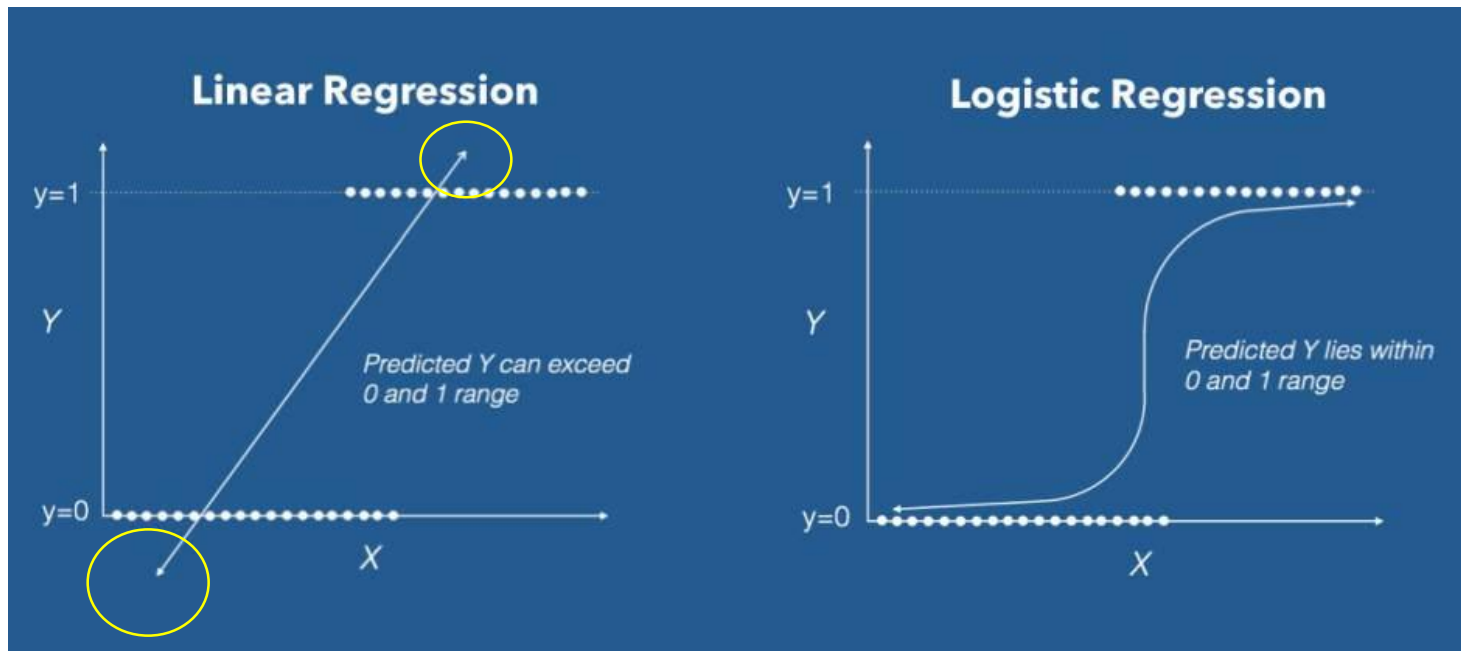
Calendário	Conteúdo	Referencial Teórico
05/02/2019	Regressão Linear (Simples e Multipla)	An Introduction to Statistical Learning with Applications in R, 2013
12/02/2019	Regressão Logística (Simples e Multipla)	
19/02/2019	Aula cancelada	
26/02/2019	Aula cancelada	Categorical Data Analysis, 2013
12/03/2019	Modelo Multinomial	
19/03/2019	Árvores de decisão e Florestas Aleatórias	An Introduction to Statistical Learning with Applications in R, 2013
26/03/2019	KNN, PCA e K Means Clustering	
08/04/2019	Avaliação	-

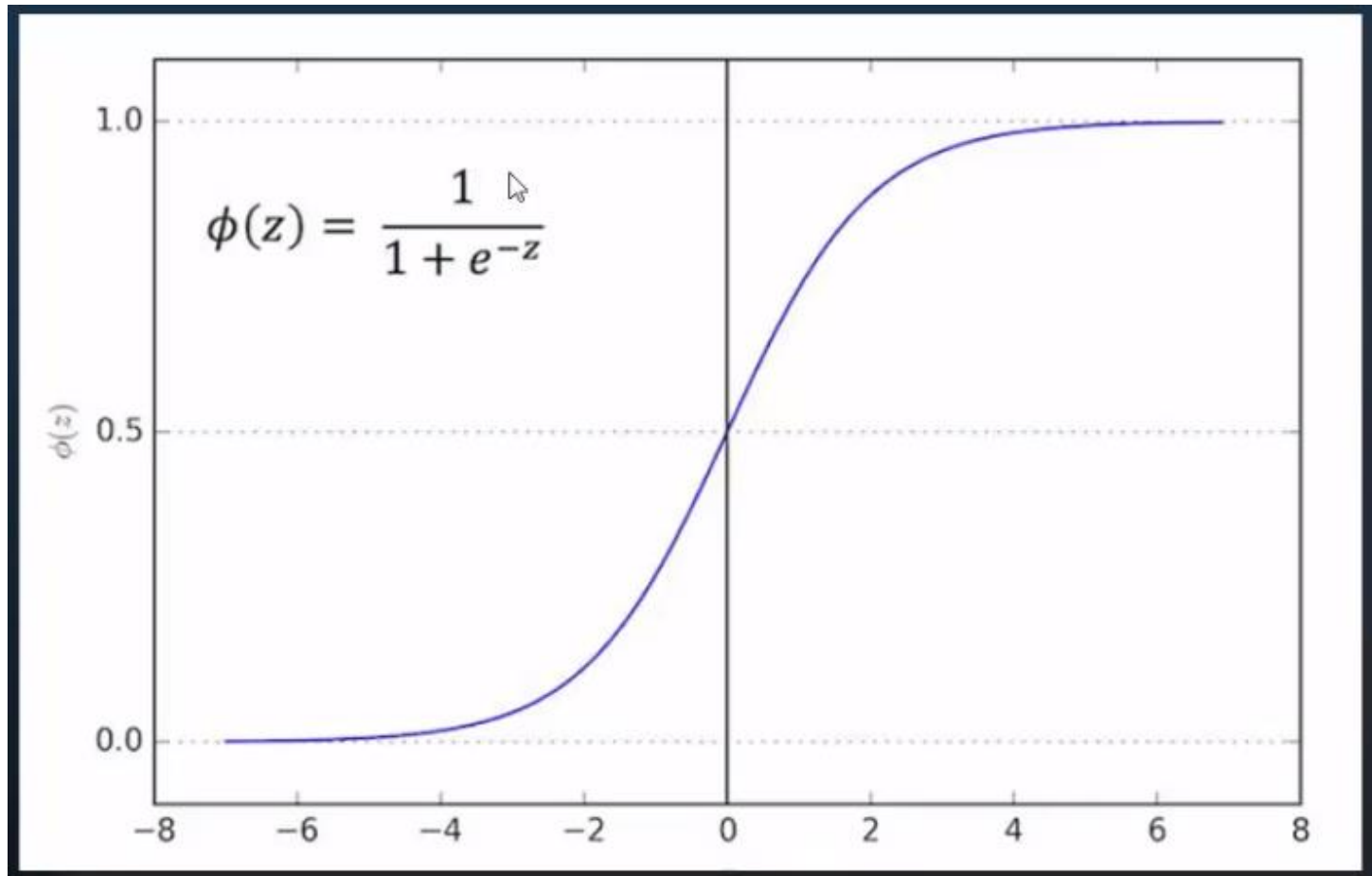
Regressão Logística Multinomial

Livro texto: *Categorical Data Analysis*

Cap. 8 – Models for Multinomial Responses

Regressão Logística





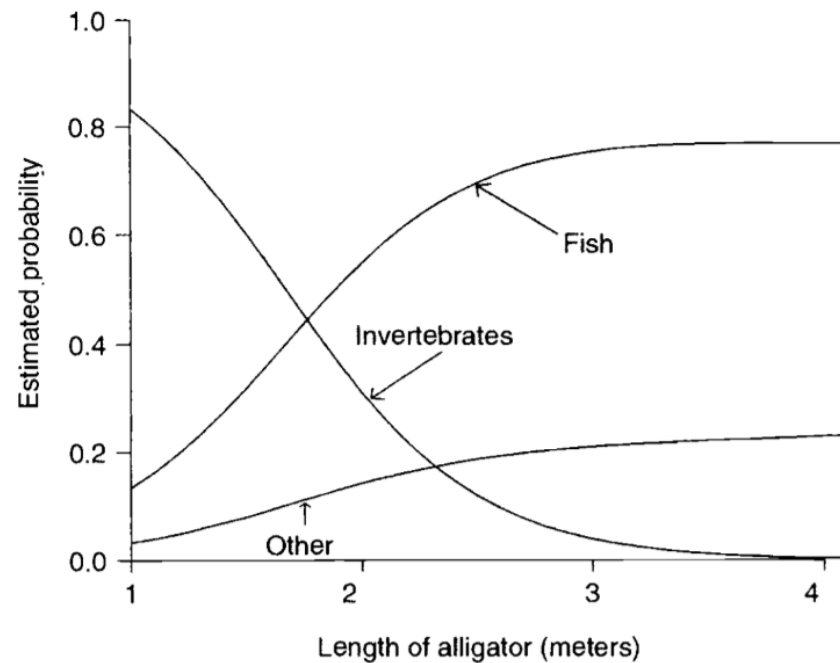


Figure 8.1 Estimated probabilities for primary food choice.

$$P(y_i = 1|x_i) = P_{i1} = \frac{1}{1 + \exp(\mathbf{X}_i\beta_{12}) + \exp(\mathbf{X}_i\beta_{13})}$$

$$P(y_i = 2|x_i) = P_{i1} = \frac{\exp(\mathbf{X}_i\beta_2)}{1 + \exp(\mathbf{X}_i\beta_{22}) + \exp(\mathbf{X}_i\beta_{23})}$$

$$P(y_i = 3|x_i) = P_{i1} = \frac{\exp(\mathbf{X}_i\beta_3)}{1 + \exp(\mathbf{X}_i\beta_{32}) + \exp(\mathbf{X}_i\beta_{33})}$$

Livro texto: *Categorical Data Analysis*

Cap. 8 – Models for Multinomial Responses

Seção 8.1.4 – Fitting Baseline-Categorical Logistic Model

Vantagens do modelo multinomial

- Facilidade para lidar com variáveis categóricas para mais de um nível
- Forte resultados em termos de probabilidade
- Facilidade de classificação de instâncias (indivíduos) em categorias
- Requer pequeno número de suposições
- Alto grau de confiabilidade

Perguntas e serem respondidas

- O modelo faz sentido?
- O modelo é útil para o objetivo pretendido? Se, por exemplo, o custo da coleta dos dados de uma variável é exorbitante e impossível de ser obtido, isso resultará em um modelo sem utilidade;
- Todos os coeficientes são razoáveis, ou seja, trazem valores que fazem sentido em termo de análise?
- A adequabilidade do modelo é satisfatória? Tem boa Precisão, Recall e F1 Score?



Interpretação dos Coeficientes

**Exercício do dia
12/02/2019**

META

- Participação em uma competição de **Machine Learning** no *Kaggle*
- Aplicar uma das técnicas de predição utilizadas no curso
- Escrever um artigo sobre modelagem preditiva no *Linkedin*
 - *Causalidade*
 - *Risco*
 - *Identificação*
 - *etc*

Exercício

Para esta aula, estaremos trabalhando com o conjunto de dados obtidos a partir do nosso livro texto. Os dados fazem referência a um estudo que tem o objetivo de identificar os fatores que influenciam a escolha alimentar dos jacarés.

Para tanto no estudo foram capturados 219 jacarés em quatro Lagos da Flórida.

A resposta para esse levantamento é nominal, retorna o tipo de alimento primário, em volume, encontrado no estômago do jacaré observado. Desta maneira temos cinco categorias:



Exercício

Links uteis

- **Abordagem em R**

<https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/>

- **Abordagem em Python**

<http://dataaspirant.com/2017/05/15/implement-multinomial-logistic-regression-python/>

Exercício

Jupyter Notebook



```
C:\Users\Pichau>conda install -c r r-essentials
```