TÓPICOS EM CIÊNCIA DE DADOS PARA O ESPORTE

ALGORITMO GLICKO

DIEGO RODRIGUES DSC

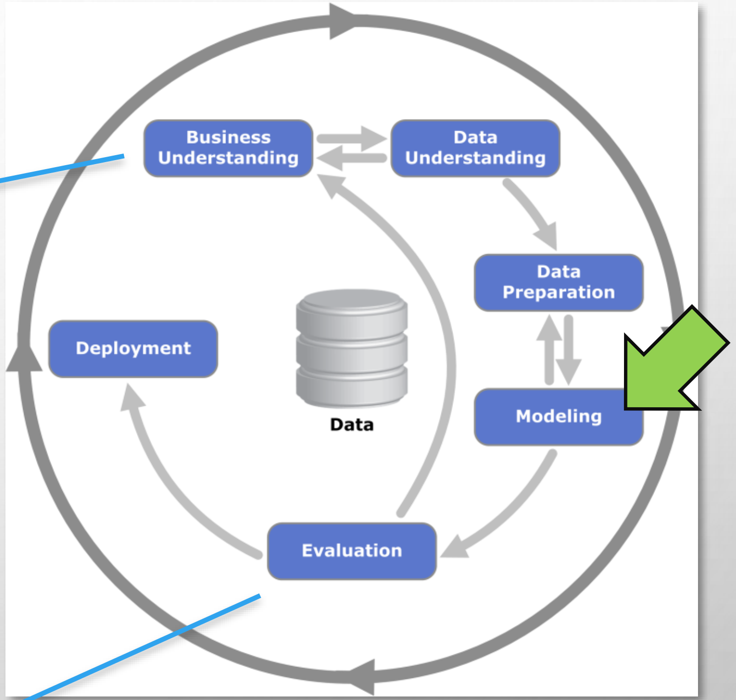
INFNET

AGENDA

- PARTE 1 : TEORIA
  - CONTEXTO
  - ALGORITMO GLICKO
- PARTE 2 : PRÁTICA
  - PROGRAMA PYTHON → RANQUEAMENTO ESTATÍSTICO DO BRASILEIRÃO (ELO VS GLICKO VS TRUESKILL)

CRONOGRAMA

DIA	NÚMERO	ÁREA	AULA	TRABALHOS
30/1/2024	1	Intro	Introdução a Disciplina e Organização do Ambiente	
1/2/2024	2	Dados	Coleta de Dados e Sensoriamento	
6/2/2024	3		Variáveis Aleatórias	Grupos
8/2/2024	4		Análise Exploratória	
15/2/2024	5		Estatísticas para Ranqueamento	
20/02/2024	6	Estatística	Ranqueamento Estatístico : ELO	
22/02/2024	7		Ranqueamento Estatístico : Glicko	
27/2/2024	8		Ranqueamento Estatístico : TrueSkill	
29/2/2024	9		Ranqueamento Estatístico : XELO	Base de Dados
5/3/2024	10		Modelos de Aprendizado de Máquina	
7/3/2024	11		Machine Learning: Classificação	
12/3/2024	12	ML	Machine Learning: Regressão	
14/3/2024	13		Machine Learning: Agrupamento	Pesquisa
19/3/2024	14		Machine Learning: Visão Computacional	
21/3/2024	15		Aplicações & Artigos: Esportes Independentes	Modelo
26/3/2024	16	Esportes	Aplicações & Artigos: Esportes de Objeto	
28/3/2024	17		Aplicações & Artigos: Esportes de Combate	
2/4/2024	18		Aplicações & Artigos : Betting	
4/4/2024	19		Workshop	
9/4/2024	20	Workshop	Apresentações de Trabalhos I	Apresentação
11/4/2024	21		Apresentações de Trabalhos II	



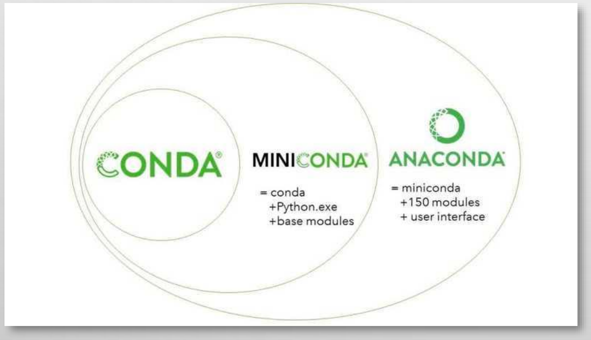
CRISP-DM

SETUP INICIAL DO AMBIENTE PYTHON

6. Estimação e Inferência

5. Visualização

4. Variáveis Aleatórias



1. Editor de Código

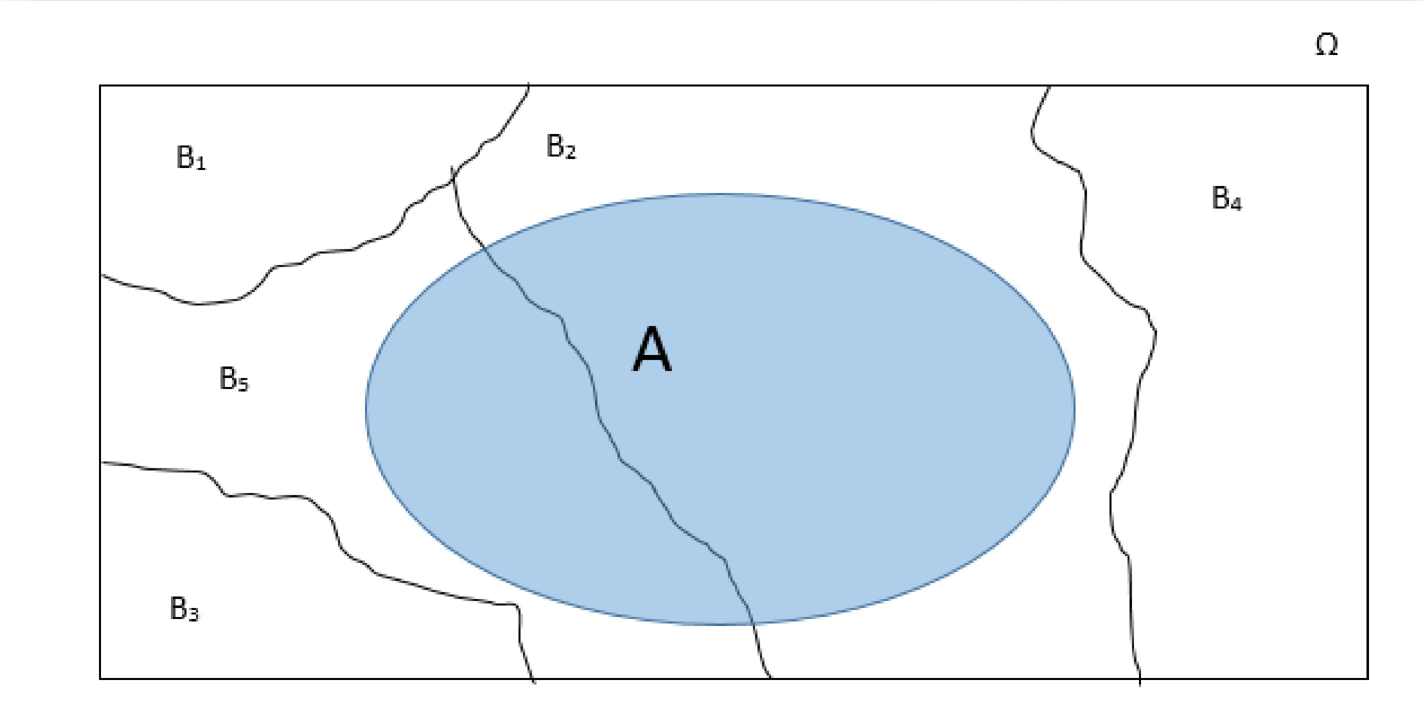
2. Gestor de Ambiente

3. Ambiente Python do Projeto

3. Notebook Dinâmico

CONTEXTO

TEOREMA DE BAYES



$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

INFERÊNCIA BAYESIANA

10.2 Bayes Rule

The distribution of a variable x conditioned on a variable y is

$$p(x \mid y) = \frac{p(x,y)}{p(y)} \tag{10.1}$$

Given that $p(y \mid x)$ can be expressed similarly we can write

$$p(x \mid y) = \frac{p(y \mid x)p(x)}{p(y)} \tag{10.2}$$

which is Baye's rule. The density $p(x)$ is known as the *prior*, $p(y \mid x)$ as the *likelihood* and $p(y)$ as the *evidence* or *marginal likelihood*. Baye's rule shows how a prior distribution can be turned into a posterior distribution ie. how we update our distribution in the light of new information. To do this it is necessary to calculate the normalising term; the evidence

$$p(y) = \int p(y \mid x)p(x)dx \tag{10.3}$$

which, being an integral, can sometimes be problematic to evaluate.

10.3 Gaussian Variables

A Gaussian random variable x has the probability density function (PDF)

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right] \tag{10.10}$$

where the mean is μ and the variance is σ^2. The inverse of the variance is known as the precision $\beta = 1/\sigma^2$. The Gaussian PDF is written in shorthand as

$$p(x) = N(x; \mu, \sigma^2) \tag{10.11}$$

If the prior is Gaussian

$$p(x) = N(x; x_0, 1/\beta_0) \tag{10.12}$$

where x_0 is the prior mean and β_0 is the prior precision and the likelihood is also Gaussian

$$p(y \mid x) = N(y; x, 1/\beta_D) \tag{10.13}$$

where the variable x is the mean of the likelihood and β_D is the data precision then the posterior distribution is also Gaussian (see eg. [33],page 37).

$$p(x \mid y) = N(x; m, 1/\beta) \tag{10.14}$$

where the mean and precision are given by

$$\beta = \beta_0 + \beta_D \tag{10.15}$$

and

$$m = \frac{\beta_0}{\beta}x_0 + \frac{\beta_D}{\beta}y \tag{10.16}$$

Thus, the posterior precision is given by the sum of the prior precision and the data precision and the posterior mean is given by the sum of the prior data mean and the new data value each weighted by their relative precisions [1].

https://www.fil.ion.ucl.ac.uk/~wpenny/

MLE VS MAP

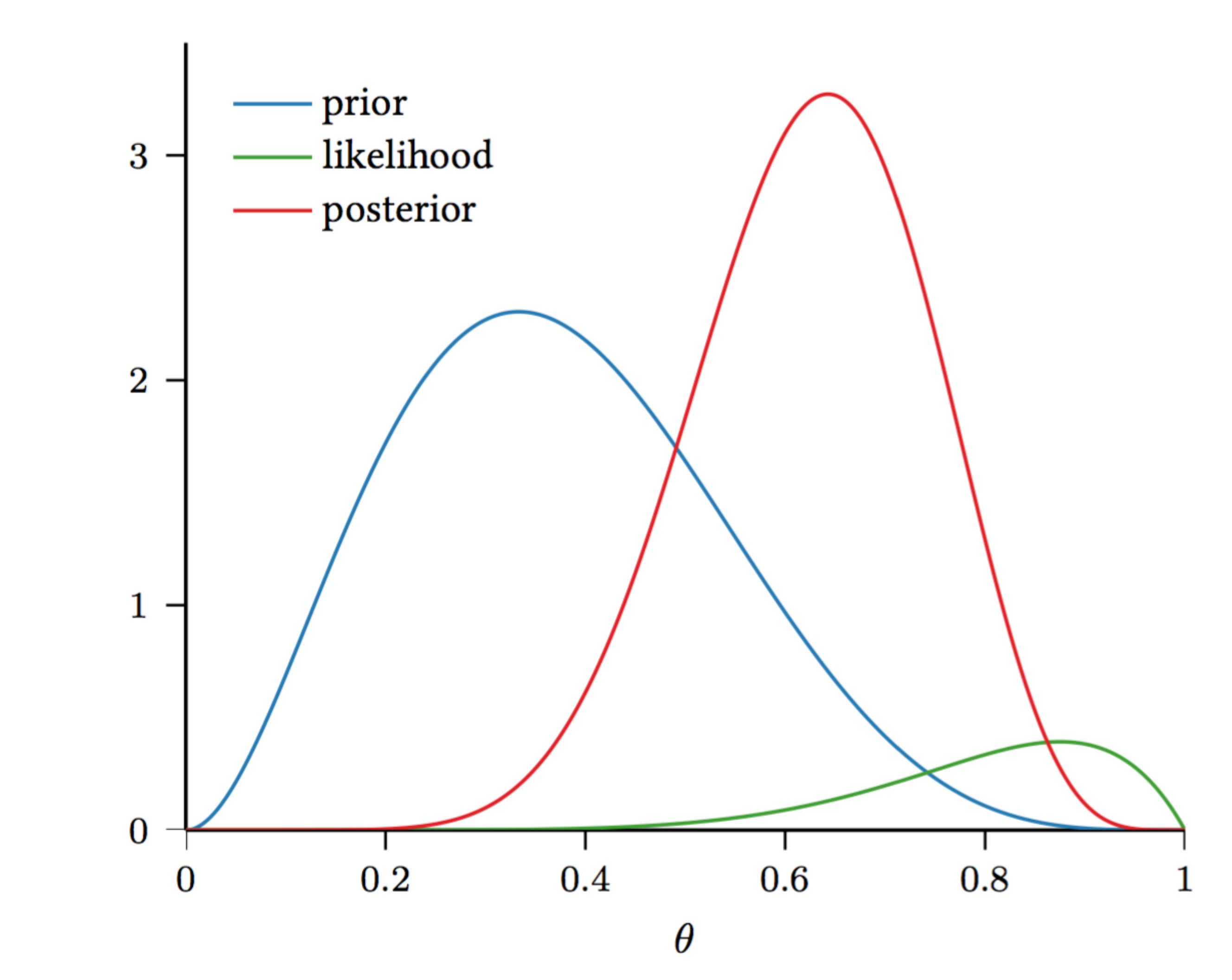
Suppose we have data $\mathcal{D} = \{x^{(i)}\}_{i=1}^{N}$

Maximum Likelihood Estimate (MLE)

$$\boldsymbol{\theta}^{\text{MLE}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta})$$

Maximum *a posteriori* (MAP) estimate

$$\boldsymbol{\theta}^{\text{MAP}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \prod_{i=1}^{N} p(\mathbf{x}^{(i)}|\boldsymbol{\theta})\underbrace{p(\boldsymbol{\theta})}_{\text{Prior}}$$



https://datascience.stackexchange.com/questions/81811/what-is-the-difference-between-maximum-likelihood-hypothesis-and-maximum-a-poste

ALGORITMO GLICKO

PAPERS ORIGINAIS

DESAFIO: RODAR O BOOTELO NA MÁQUINA II

PRÓXIMA AULA
LEITURA: ALGORITMO TRUESKILL