



FIAP



MACHINE LEARNING & MODELLING

Esta disciplina aborda os principais conceitos sobre aprendizado de máquina e as técnicas clássicas de modelagem



Na última aula...

- Quis: problemas X tipos de modelagem
- Forças & Fraquezas de ML
- Pipeline de Data Science na prática
- Desafios de trabalhar com dados (*overfitting, underfitting*)

Agenda

- Conceitos avançados sobre dados
- Árvores de decisão

Tipos de dados

Vamos aprender a reconhecer os tipos de dados existentes

Classificação dos dados

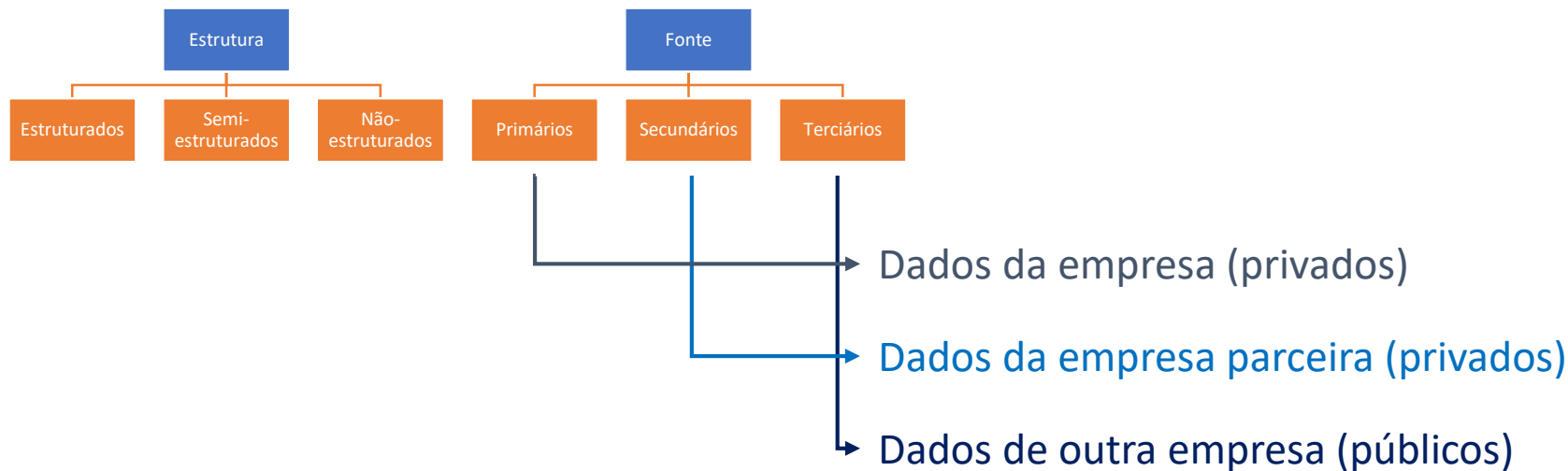
JSON



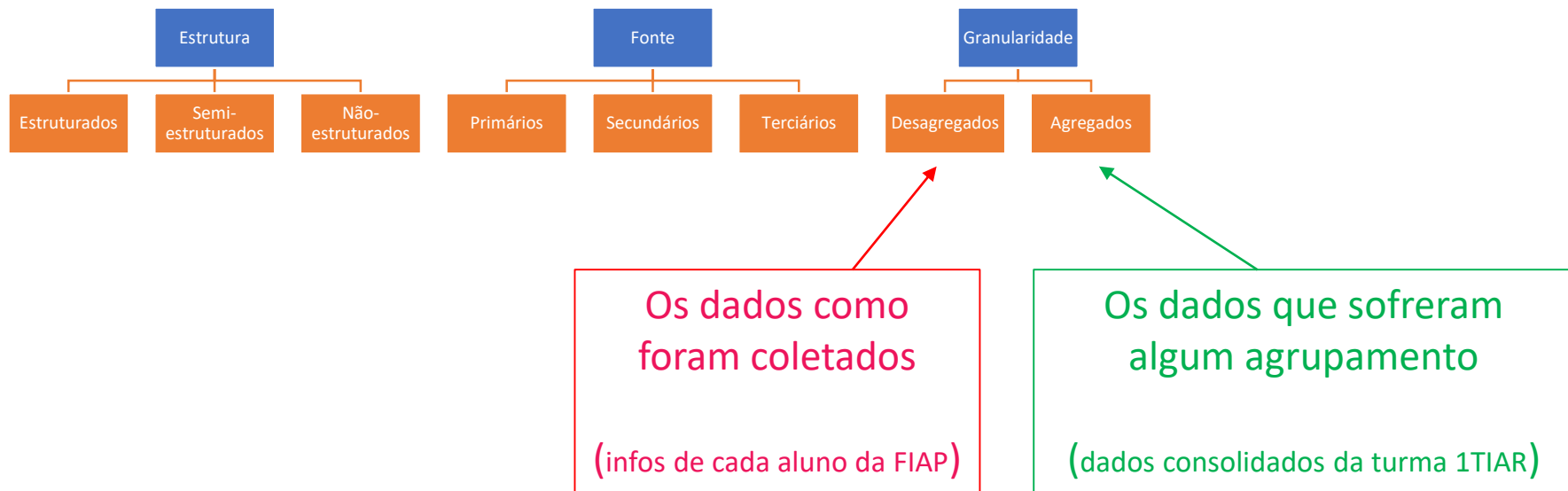
```
{
  "image": {
    "width": 800,
    "height": 600,
    "title": "JSONlogo",
    "thumbnail": {
      "url": "http://www.json.org/img/json160.gif",
      "height": 160,
      "width": 160
    },
  },
  "animated": false,
  "IDs": [116, 943,
234]
}
```

```
{
  bairro: "Farolândia",
  cidade: "Aracaju",
  cep: "49032490",
  logradouro: "Avenida Murilo Dantas",
  - estado_info: {
    area_km2: "21.918,493",
    codigo_ibge: "28",
    nome: "Sergipe"
  },
  - cidade_info: {
    area_km2: "181,857",
    codigo_ibge: "2800308"
  },
  estado: "SE"
}
```

Classificação dos dados

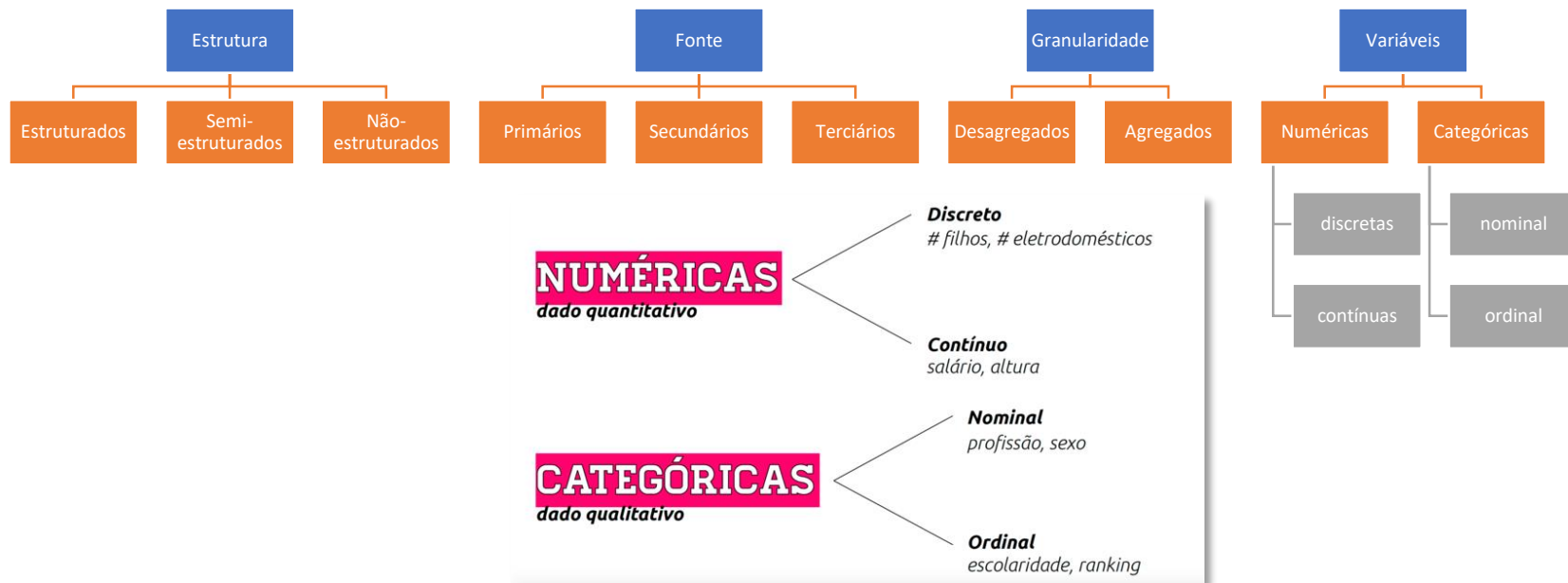


Classificação dos dados



Classificação dos dados

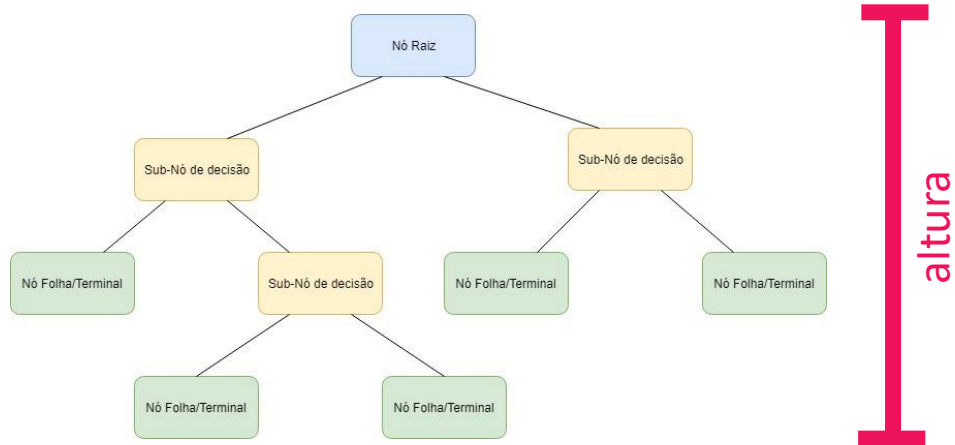
Vamos explorar estas classificações quando usarmos um dataset!



Nosso primeiro modelo

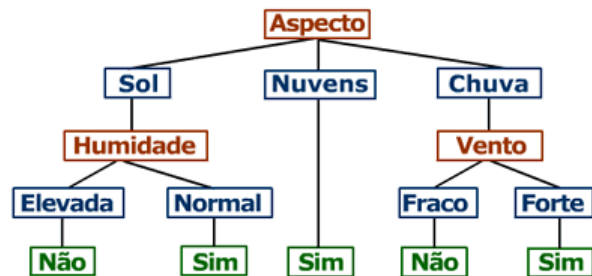
Aprendizado supervisionado >> Árvores de decisão

Árvores de decisão: terminologia



É um diagrama de fluxo

Árvore de Decisão para Jogar Ténis



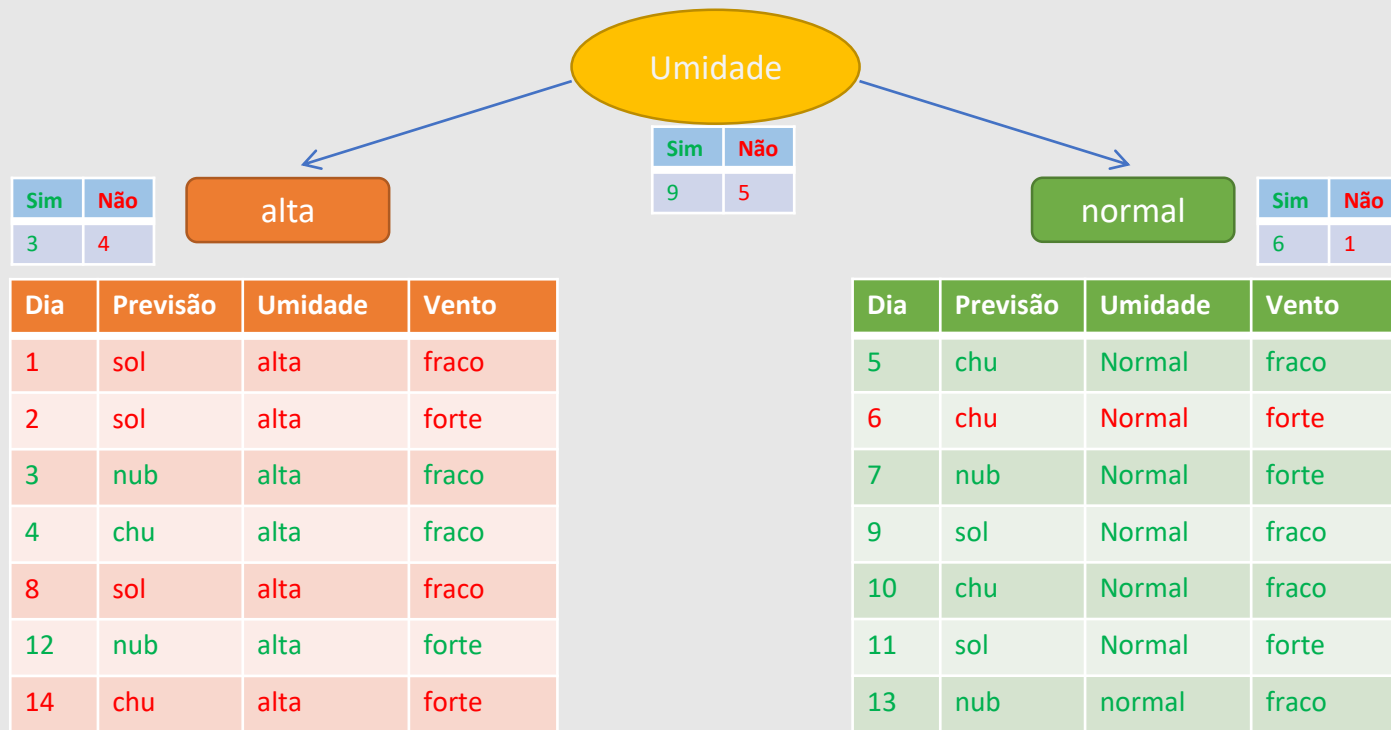
Árvores de decisão - Exemplo

Dia	Previsão	Umidade	Vento	Jogou?
1	ensolarado (sol)	alta	fraco	não
2	ensolarado (sol)	alta	forte	não
3	nublado (nub)	alta	fraco	sim
4	chuvoso (chu)	alta	fraco	sim
5	chuvoso (chu)	normal	fraco	sim
6	chuvoso (chu)	normal	forte	não
7	nublado (nub)	normal	forte	sim
8	ensolarado (sol)	alta	fraco	não
9	ensolarado (sol)	normal	fraco	sim
10	chuvoso (chu)	normal	fraco	sim
11	ensolarado (sol)	normal	forte	sim
12	nublado (nub)	alta	forte	sim
13	nublado (nub)	normal	fraco	sim
14	chuvoso (chu)	alta	forte	não
15	chuvoso (chu)	alta	fraco	??????

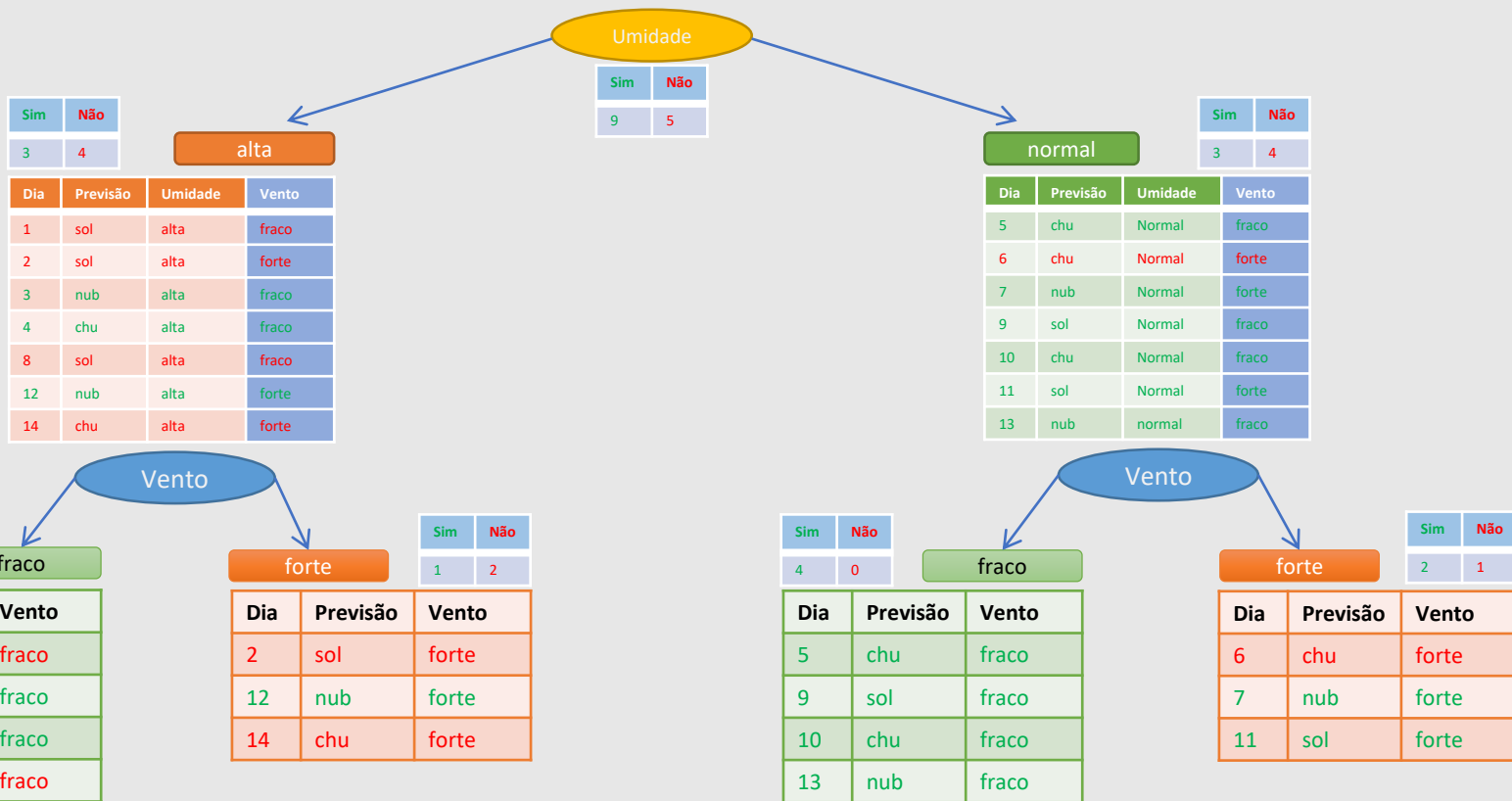
Árvores de decisão - Exemplo

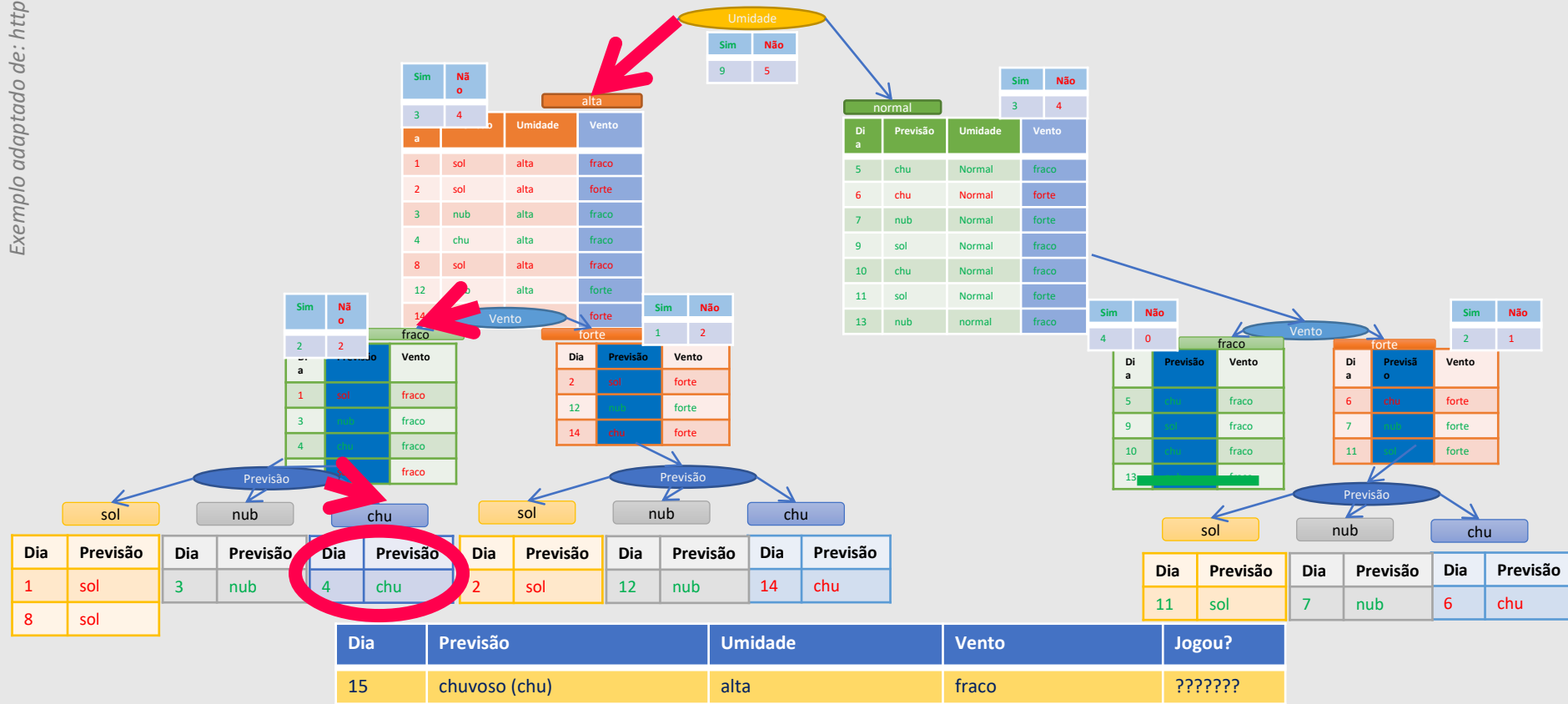
Dia	Previsão	Umidade	Vento	Jogou?
1	ensolarado (sol)	alta	fraco	não
2	ensolarado (sol)	alta	forte	não
3	nublado (nub)	alta	fraco	sim
4	chuvoso (chu)	alta	fraco	sim
5	chuvoso (chu)	normal	fraco	sim
6	chuvoso (chu)	normal	forte	não
7	nublado (nub)	normal	forte	sim
8	ensolarado (sol)	alta	fraco	não
9	ensolarado (sol)	normal	fraco	sim
10	chuvoso (chu)	normal	fraco	sim
11	ensolarado (sol)	normal	forte	sim
12	nublado (nub)	alta	forte	sim
13	nublado (nub)	normal	fraco	sim
14	chuvoso (chu)	alta	forte	não
15	chuvoso (chu)	alta	fraco	??????

Árvores de decisão - Exemplo



Árvores de decisão - Exemplo



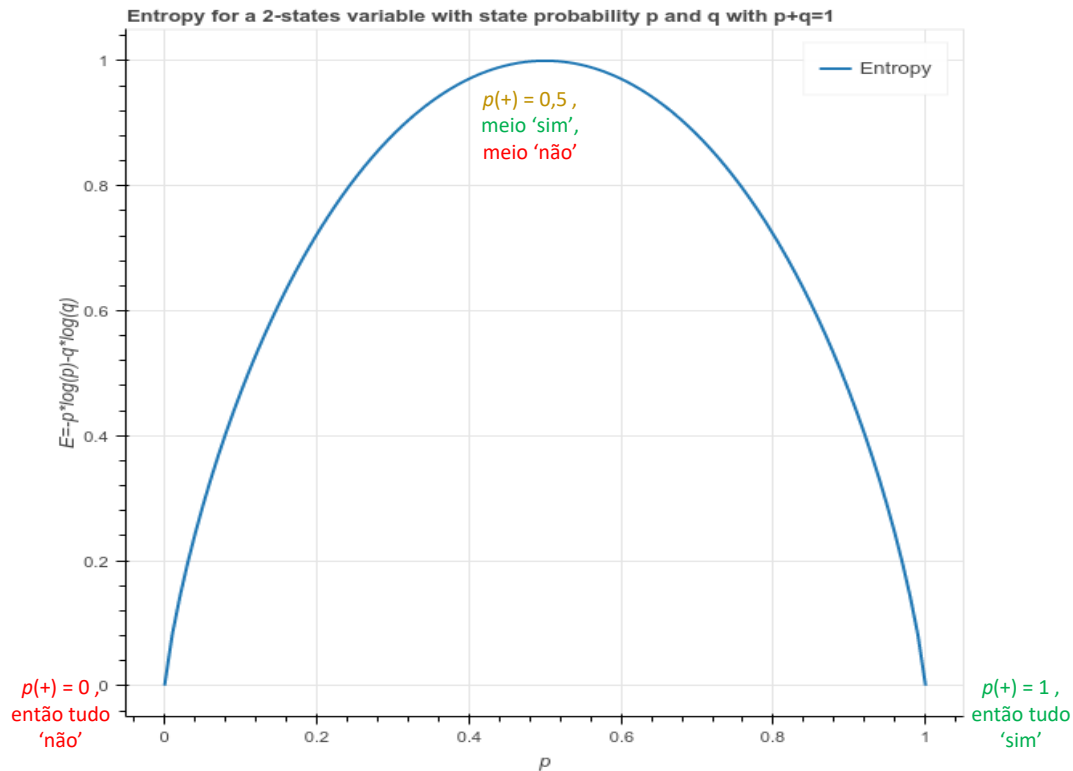


Árvores de decisão

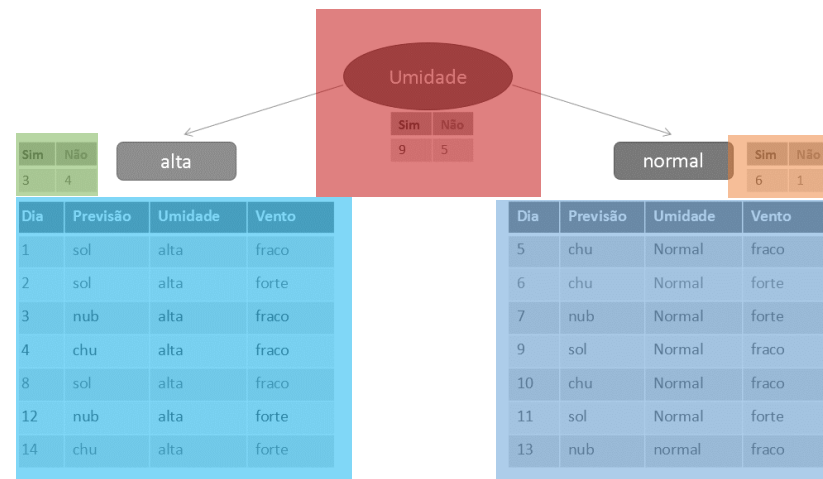
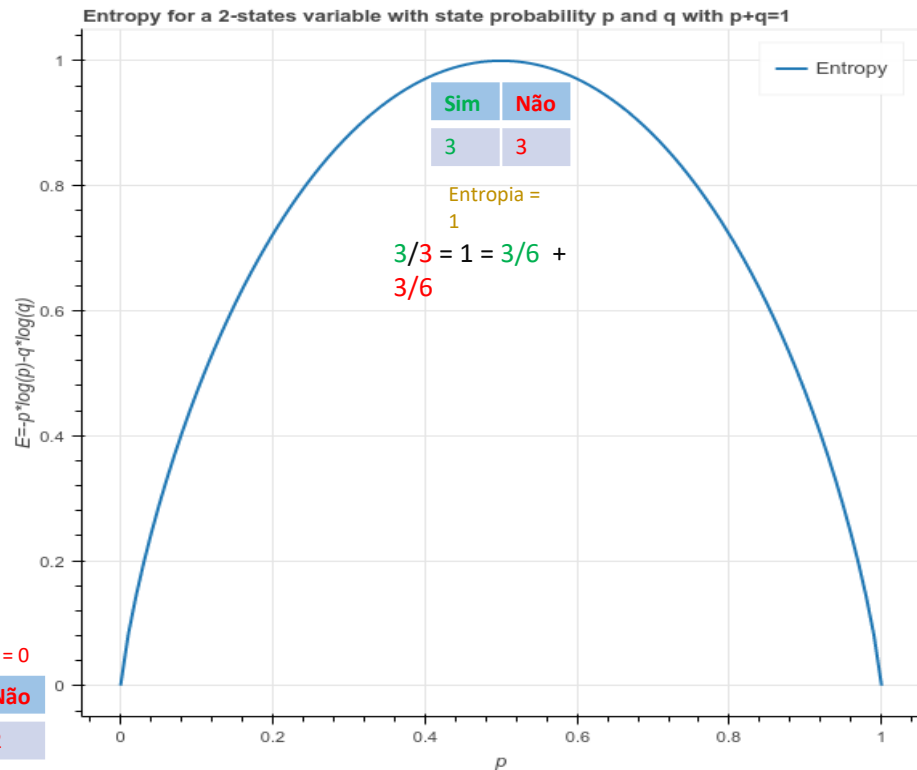
- Só existe esta árvore?
- Como podemos fazer uma árvore **melhor**?



Árvores de decisão — Entropia



Árvores de decisão – Entropia & Ganho de Informação



$$\text{Entropia(NÓ)} = - [p_1 \cdot \log(p_1) + p_2 \cdot \log(p_2) \dots + p_n \cdot \log(p_n)]$$

$$GI = \text{entropia(umidade)}$$

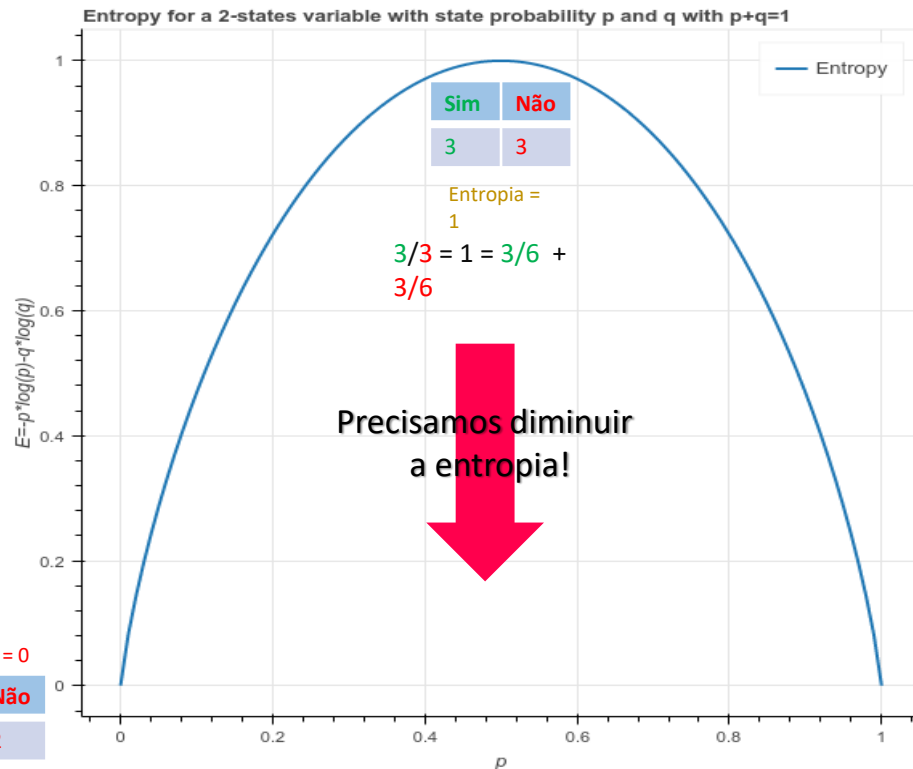
$$- p(\text{umidade=alta}) \cdot \text{entropia(u=alta)}$$

$$- p(\text{umidade=normal}) \cdot \text{entropia(u=normal)}$$

Entropia = 0

Sim	Não
4	0

Árvores de decisão – Entropia & Ganho de Informação



- Precisamos de conjuntos mais puros
- Precisamos diminuir a entropia
- Diminuindo a entropia, aumentamos o ganho de informação!

Árvores de decisão

Vantagens

- Pouco esforço de preparação dos dados
- São capazes de lidar com múltiplos labels (classes)
- Alta explicabilidade

Desvantagens

- Podem ser desbalanceadas (*overfitting*)
- São instáveis (alta variância)
- Podem não ser ótimas

Como contornar os 2 últimos problemas?



E na prática?

Vamos ver como criar um **modelo preditivo** baseado em árvore de decisão!

OBRIGADO!



Prof. Michel Fornaciali

<https://www.linkedin.com/in/michelfornaciali/>
profmichel.fornaciali@fiap.com.br

Copyright © 2021 | Professor Michel Fornaciali

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor.