

Classificador Ingênuo De Bayes

April 27, 2021

Overview I

1 Base de Dados

- Variáveis de Entrada
- Variável de Saída

2 Análise Exploratória dos Dados

- Descrição Estatística dos Dados
- Gráficos

3 Classificador Ingênuo de Bayes

- História

4 Classificador Ingênuo de Bayes

- Definição Formal do Teorema de Bayes
- Tipos de Classificadores Ingênuo de Bayes
- Vantagens
- Desvantagens
- Sobre o Projeto

5 Experimentos

- Experimentos Iniciais

Overview II

- Usando Apenas a Variável Age para Treino
- Usando Apenas Variáveis Numéricas Para Treino

6 Análise dos Resultados

- Resultados Iniciais
- Perfil Mais Receptivo
- Perfil Menos Receptivo
- Categorias exóticas

Base de Dados

Variáveis de Entrada

- age
- job
- marital
- education
- default
- balance
- housing
- loan
- contact
- day
- duration
- month
- campaign
- pdays
- previous
- poutcome

Base de Dados

Variável de Saída

- y

Análise Exploratória dos Dados

Descrição Estatística dos Dados

- count
- unique
- top
- freq
- mean
- std
- min
- 25%
- 50%
- 75%
- max

Análise Exploratória dos Dados

Gráficos

Figure: Exemplo

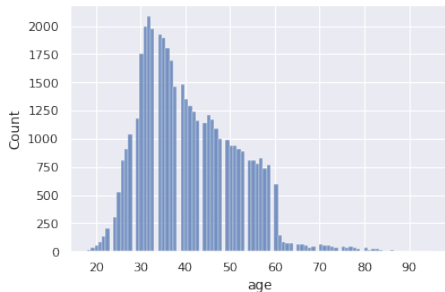
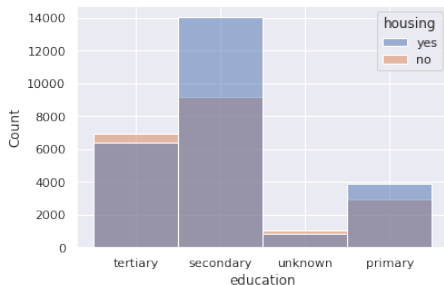


Figure: Exemplo2



Análise Exploratória dos Dados

Gráficos

Figure: Exemplo 3

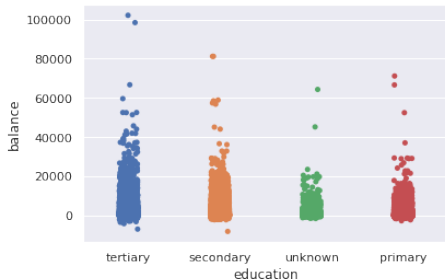
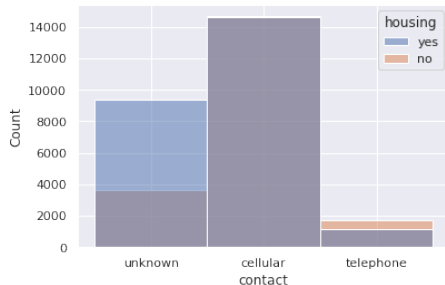


Figure: Exemplo 4



Análise Exploratória dos Dados

Gráficos

Figure: Exemplo 5

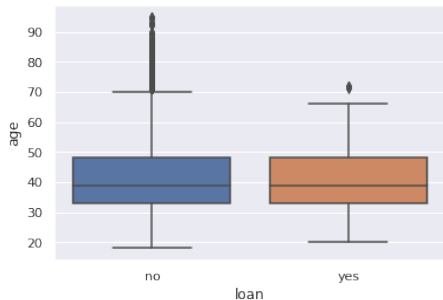
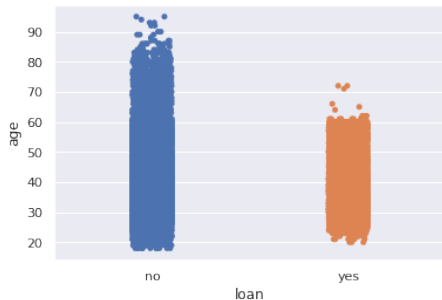


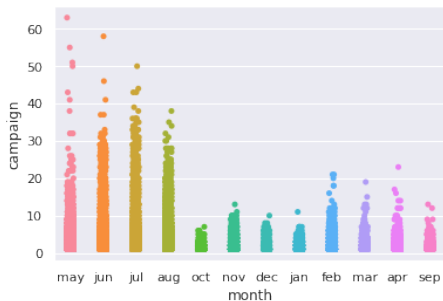
Figure: Exemplo 6



Análise Exploratória dos Dados

Gráficos

Figure: Exemplo 7



Classificador Ingênuo de Bayes

O que é Naive Bayes

Baseado no Teorema de Bayes, nome em homenagem ao matemático e pastor presbiteriano inglês Thomas Bayes, que formulou uma função probabilística com o ideal de provar a existência de Deus, é um algoritmo de classificação probabilística muito utilizado para aprendizado de máquina (Machine Learning).

Classificador Ingênuo de Bayes

Definição Formal do Teorema de Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

- $P(A|B)$: Probabilidade do evento A ocorrer dado que o evento B ocorreu.
- $P(B|A)$: Probabilidade do evento B ocorrer dado que o evento A ocorreu.
- $P(A)$: Probabilidade do evento A ocorrer
- $P(B)$: Probabilidade do evento B ocorrer.

Classificador Ingênuo de Bayes

Tipos de Classificadores Ingênuo de Bayes

- Bayes Ingênuo Gaussiano
- Bayes Ingênuo Categórico

Classificador Ingênuo de Bayes

Vantagens

- Rápido
- Eficiente
- Lida com múltiplos tipos de dados
- Ignora características irrelevantes

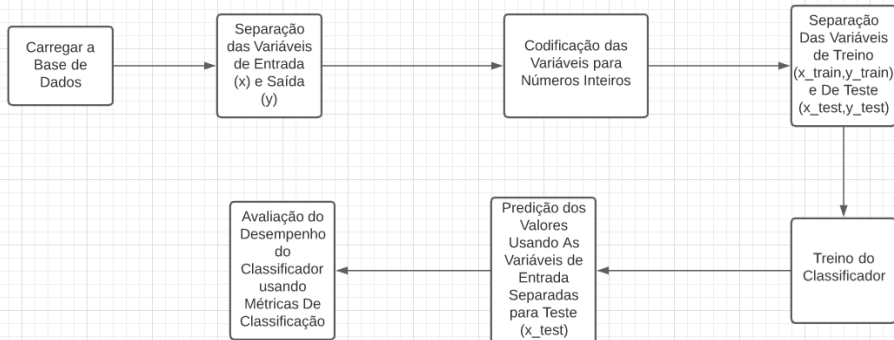
Classificador Ingênuo de Bayes

Desvantagens

- Previsão falha em frequência zero
- Ignorar a correlação entre as variáveis

Classificador Ingênuo de Bayes

Sobre o Projeto



- Precision:

$$\frac{t_p}{t_p + f_p} \quad (2)$$

- Accuracy

$$\frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (3)$$

- Recall Score

$$\frac{t_p}{t_p + f_n} \quad (4)$$

- F1 Score

$$\frac{2 \cdot (\textit{precision} \cdot \textit{recall})}{\textit{precision} + \textit{recall}} \quad (5)$$

Experimentos

Experimentos Iniciais

	Categórico	Gaussiano
Precision	0.89	0.84
Accuracy	0.89	0.84
Recall Score	0.89	0.84
F1-Score	0.89	0.84

Table: Relatório de Classificação por Label do Classificador Categórico.

	0	1
Precision	0.93	0.53
Recall Score	0.95	0.43
F1-Score	0.94	0.47

Table: Relatório de Classificação por Label do Classificador Gaussiano

	0	1
Precision	0.89	0.49
Recall Score	1.00	0.03
F1-Score	0.94	0.06

Experimentos

Experimentos Iniciais

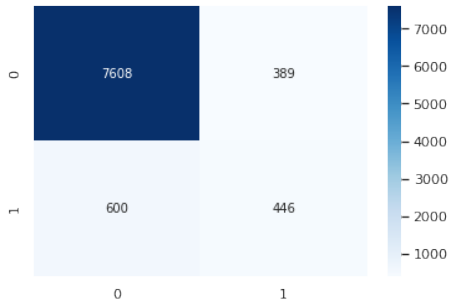


Figure: Confusion Matrix do Classificador Categórico

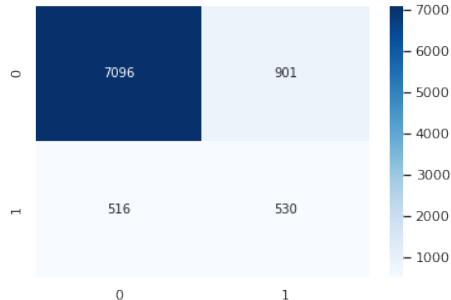


Figure: Confusion Matrix do Classificador Gaussiano

Experimentos

Usando Apenas a Variável Age para Treino

	Categórico	Gaussiano
Precision	0.88	0.88
Accuracy	0.88	0.88
Recall Score	0.88	0.88
F1-Score	0.88	0.88

Table: Relatório de Classificação por Label do Classificador Categórico.

	0	1
Precision	0.88	0.50
Recall Score	1.00	0.02
F1-Score	0.94	0.04

Table: Relatório de Classificação por Label do Classificador Gaussiano

	0	1
Precision	0.88	0.48
Recall Score	1.00	0.03
F1-Score	0.94	0.05

Experimentos

Usando Apenas a Variável Age para Treino

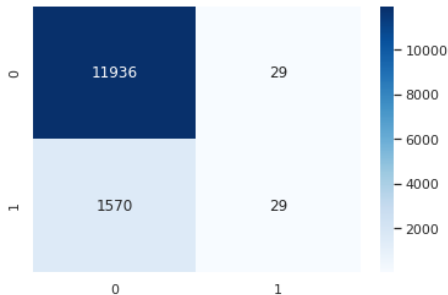


Figure: Confusion Matrix do Classificador Categórico

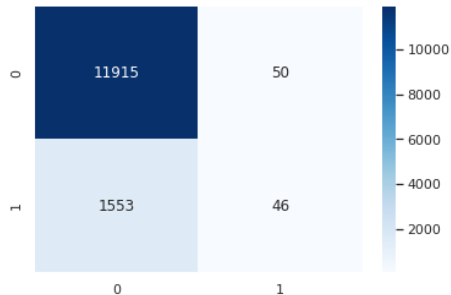


Figure: Confusion Matrix do Classificador Gaussiano

Experimentos

Usando Apenas Variáveis Numéricas Para Treino

Table: Comparação entre o Classificador Categórico e o Gaussiano

	Categórico	Gaussiano
Precision	0.89	0.89
Accuracy	0.89	0.89
Recall Score	0.89	0.89
F1-Score	0.89	0.89

Table: Relatório de Classificação por Label do Classificador Categórico.

	0	1
Precision	0.89	0.63
Recall Score	0.99	0.11
F1-Score	0.94	0.18

Table: Relatório de Classificação por Label do Classificador Gaussiano

	0	1
Precision	0.91	0.53
Recall Score	0.96	0.32
F1-Score	0.94	0.40

Experimentos

Usando Apenas Variáveis Numéricas Para Treino

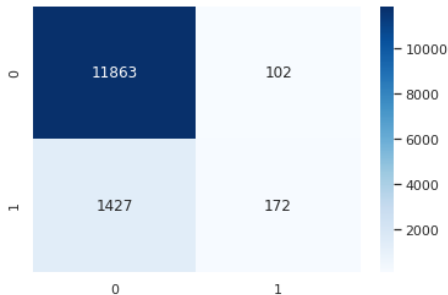


Figure: Confusion Matrix do Classificador Categórico

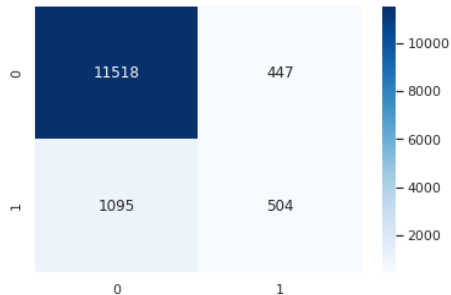


Figure: Confusion Matrix do Classificador Gaussiano

Análise dos Resultados

Resultados Iniciais

- Classificador Gaussiano vs Classificador Categórico
- Número de Features
- Falsos positivos vs Falsos Negativos

Análise dos Resultados

Perfil Mais Receptivo

- Profissão: estudante
- Estado Civil: divorciado
- Credito Pessoal: possui
- Credito de Habitação: não possui
- Tipo de Contato: celular
- Educação: ensino médio
- Mês da Campanha: setembro

Análise dos Resultados

Perfil Menos Receptivo

- Profissão: operario
- Estado Civil: casado
- Credito Pessoal: não possui
- Credito de Habitação: possui
- Tipo de Contato: unknown
- Educação: ensino fundamental
- Mês da Campanha: maio

Análise dos Resultados

Categorias exóticas

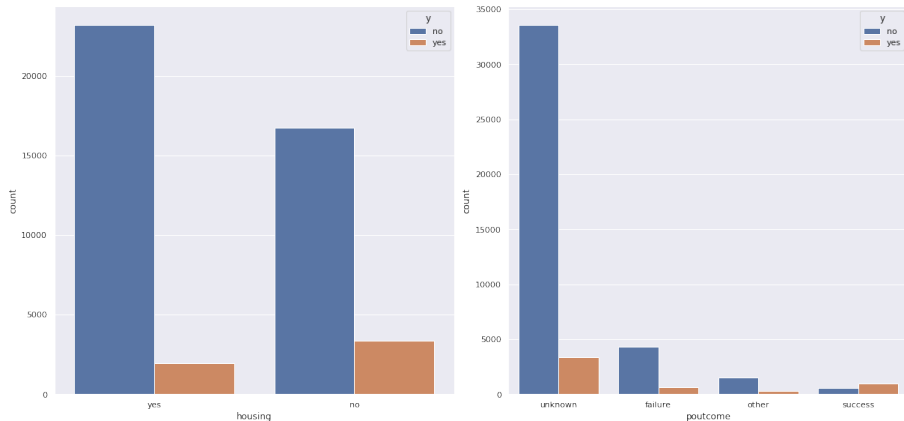


Figure: Categorias mais exóticas