

MACHINE LEARNING COM AMAZON AWS E SAGEMAKER



CONTEÚDO DO CURSO

1. Introdução ao AWS

- AWS, S3, EC2, IAM e SageMaker
- Integração do SageMaker com o S3

2. Regressão com Linear Learner e XGBoost

- Previsão do preço de casas e limite do cartão de crédito
- Tuning

3. Classificação com Linear Learner e XGBoost

- Base de dados do censo

CONTEÚDO DO CURSO

4. Séries temporais com DeepAR

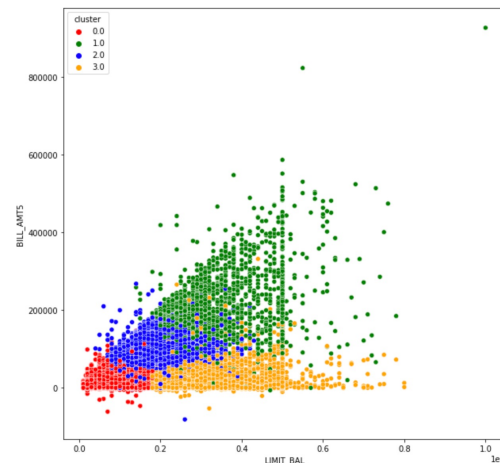
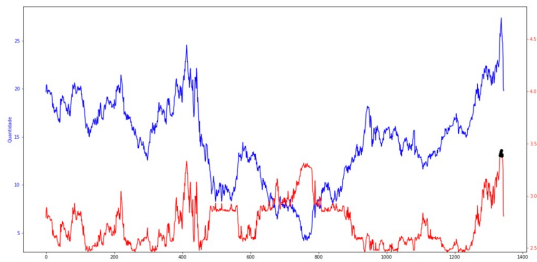
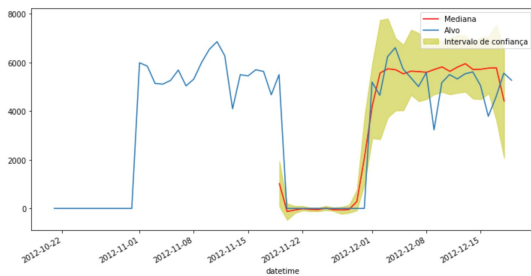
- Base de dados de aluguel de bicicletas e preço das ações

5. Outliers com Random Cut Forest

- Base de dados de aluguel de bicicletas e preço das ações

6. PCA e agrupamento com k-means

- PCA + agrupamento (base cartão de crédito)
- PCA + classificação (base do censo)



CONTEÚDO DO CURSO

7. Redes neurais para classificação de imagens

- Base de dados Caltech 256

8. Integração do SageMaker com o TensorFlow

- Debug

9. Endpoint externo

10. Autopilot (aprendizagem automática)

11. ANEXOS

- Redes neurais artificiais
- Redes neurais convolucionais
- Redes neurais recorrentes

INFORMAÇÕES ADICIONAIS

1. Pré-requisitos: lógica de programação, programação básica em Python e conceitos básicos de Machine Learning
2. Intuição básica sobre os algoritmos
3. Foco em como utilizar o **SageMaker**
4. Não é um curso sobre certificação!

COMPUTAÇÃO EM NUVEM

- Acesso a serviços na Internet com pagamento de acordo com o uso



AWS – AMAZON WEB SERVICES

- Integração de aplicativos
- Blockchain
- Computação
- Banco de dados
- IoT
- Machine learning
- Redes
- Segurança
- Armazenamento
- Mais serviços: <https://aws.amazon.com/pt/solutions/>



Fonte: <https://aws.amazon.com/about-aws/global-infrastructure/>

AMAZON S3 – SIMPLE STORAGE SERVICE



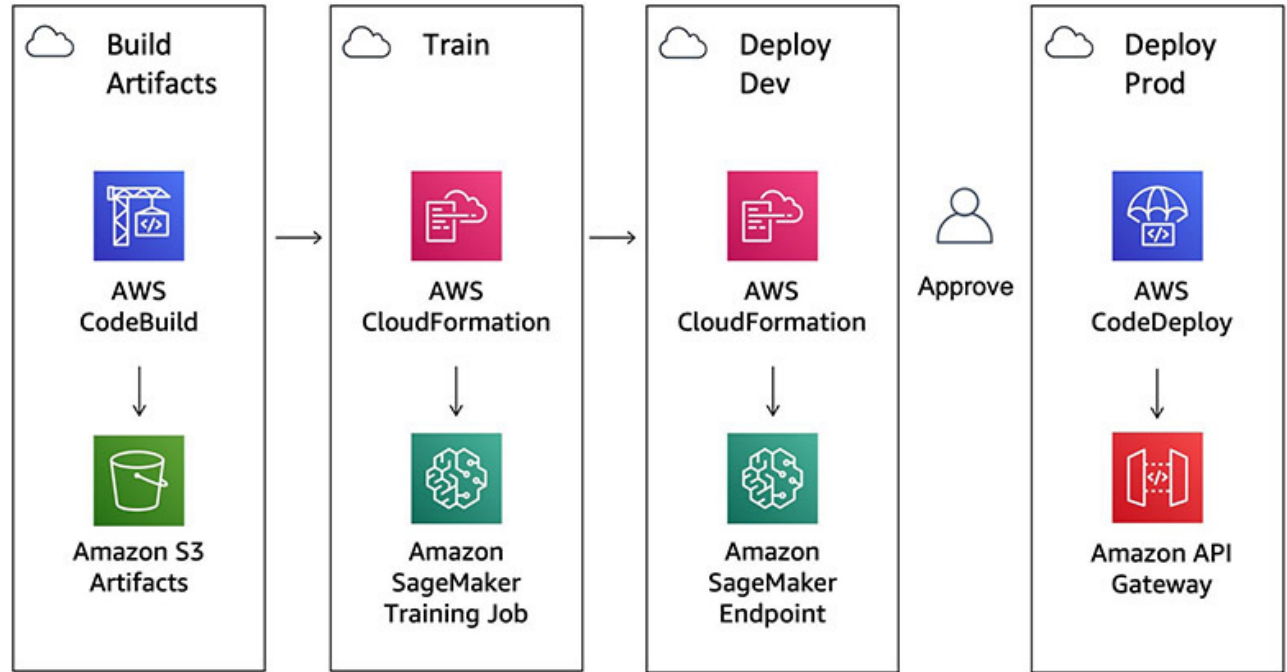
AMAZON EC2 – ELASTIC COMPUTING CLOUD



AMAZON IAM – IDENTITY AND ACCESS MANAGEMENT



AMAZON SAGEMAKER

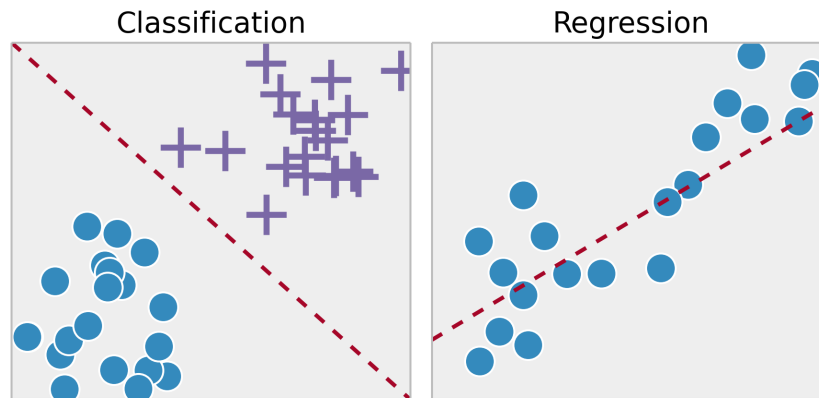


REGRESSÃO

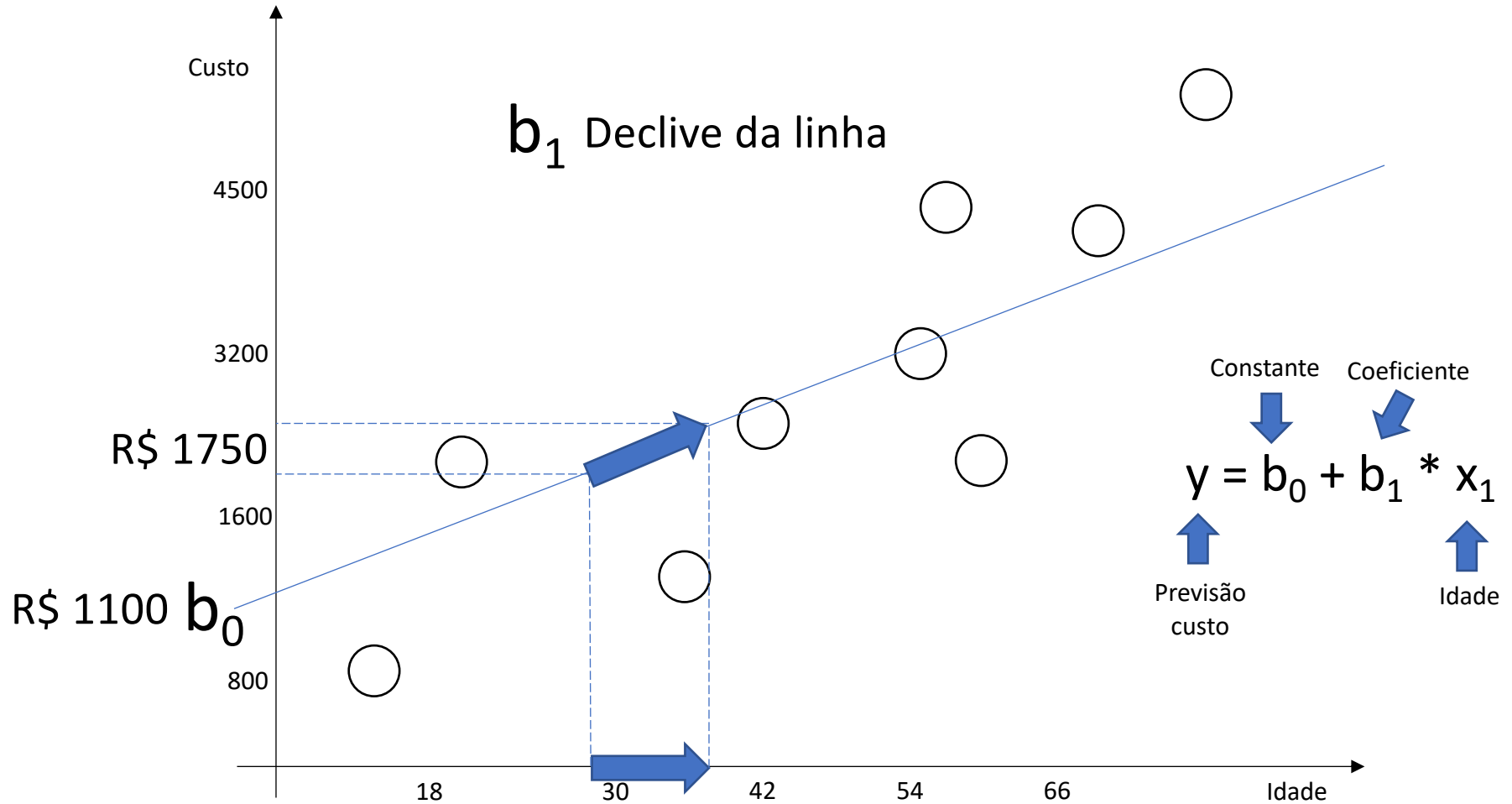
1. Base de dados de casas
2. Configurações do SageMaker
3. Linear learner (regressão linear)
4. Deploy do modelo
5. XGBoost
6. Tuning do algoritmo
7. Exercício

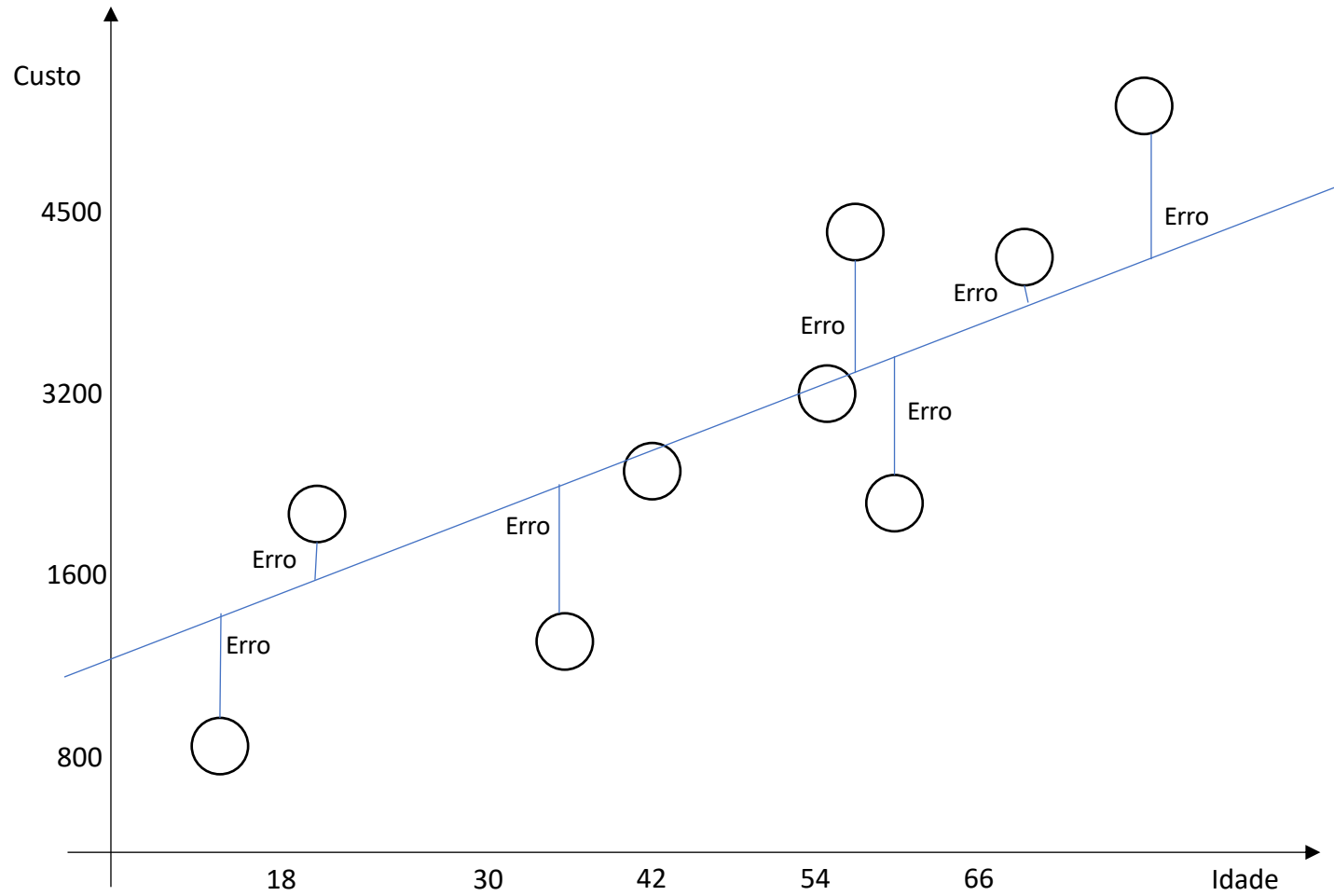
REGRESSÃO LINEAR

- Modelagem da relação entre variáveis numéricas (variável dependente y e variáveis explanatórias x)
- Temperatura, umidade e pressão do ar (x) \rightarrow velocidade do vento (y)
- Gastos no cartão de crédito, histórico (x) \rightarrow limite do cartão (y)
- Idade (x) \rightarrow custo plano de saúde (y)
- Tamanho da casa (x) \rightarrow preço da casa (y)



Relação linear entre os atributos: quanto maior a idade, maior o custo
 b_0 e b_1 definem a localização da linha (treinamento)





MEAN SQUARED ERROR

$$MSE = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2$$

Preço real	Preço calculado	Erro
150	180	$(150 - 180)^2 = 900$
60	55	$(60 - 55)^2 = 25$
220	230	$(220 - 230)^2 = 100$
45	67	$(45 - 67)^2 = 484$

Soma = 1.509

$MSE = 1.509 / 4 = 377,25$

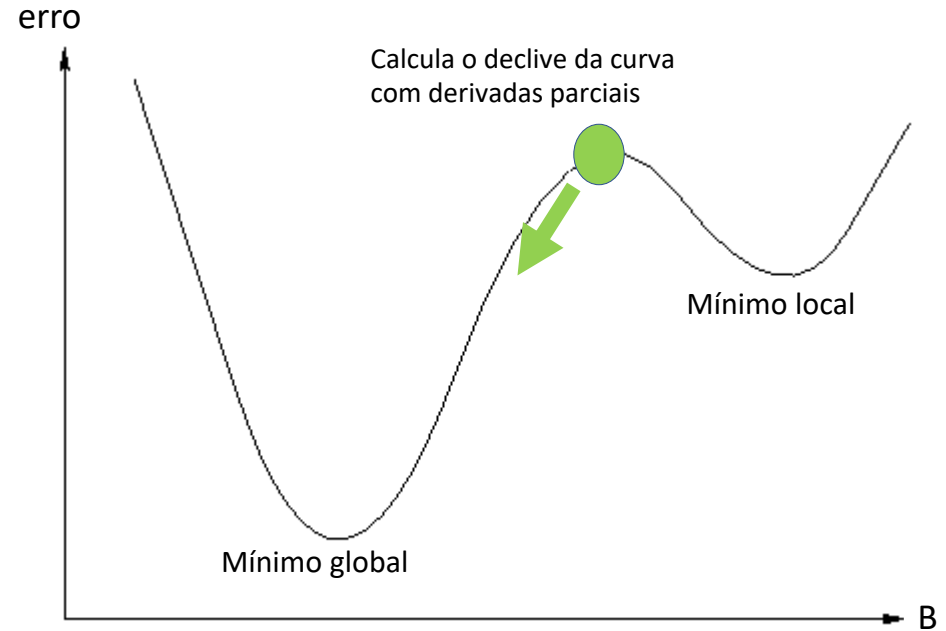
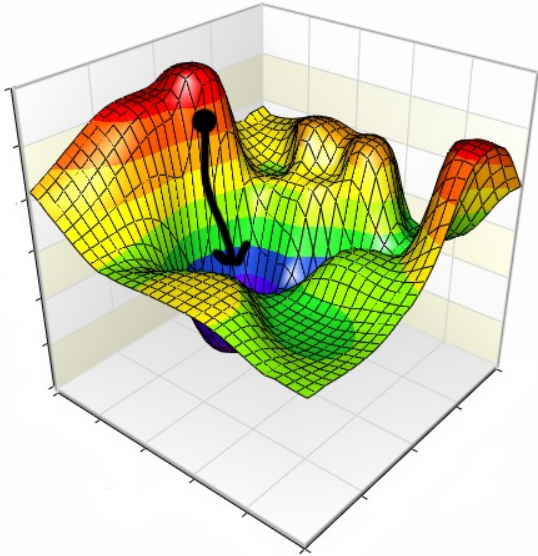
$$y = b_0 + b_1 * x_1$$

Objetivo: ajustar os parâmetros b_0 e b_1 para ter o menor erro!

DESCIDA DO GRADIENTE

$\min C(B_1, B_2 \dots B_n)$

Taxa de aprendizagem



REGRESSÃO LINEAR MÚLTIPLA

$$y = b_0 + b_1 * x_1$$

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + ... + b_n * x_n$$

XGBOOST (EXTREME GRADIENT BOOSTING)

Árvore de decisão



Fonte: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/>

Random Forest



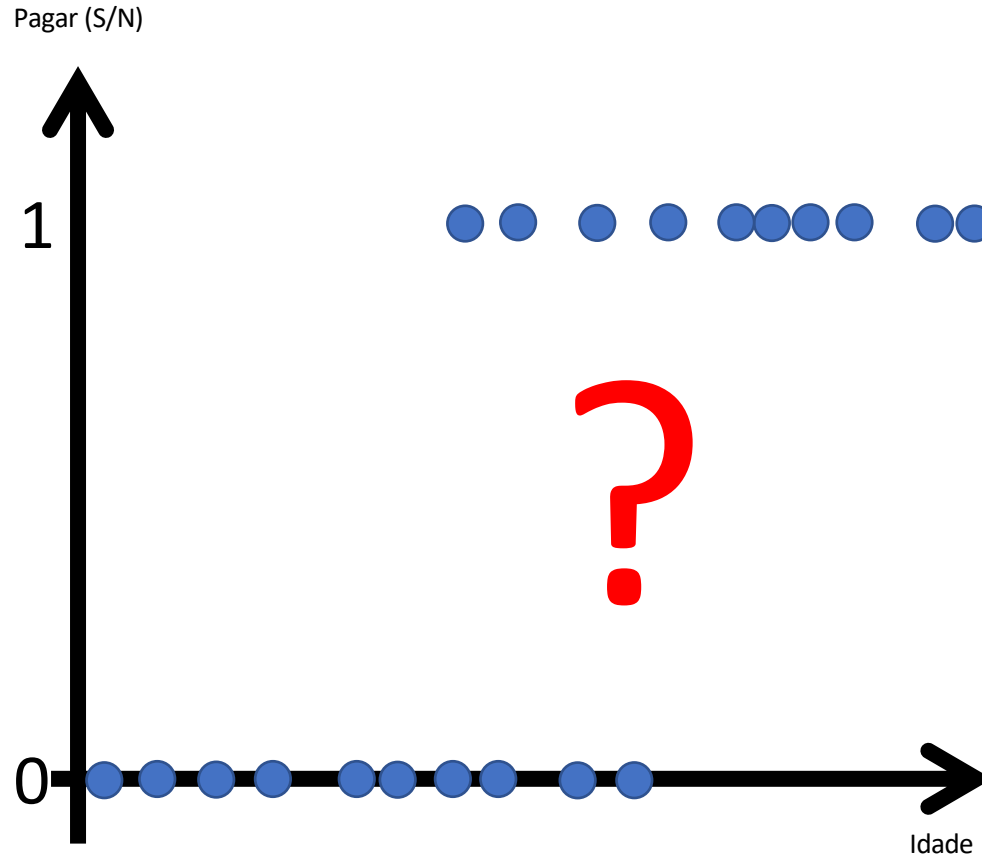
Árvore de decisão
Bagging
Random forest
Boosting
Gradient boosting
XGBoost

Artigo: <https://iaexpert.academy/2019/04/18/xgboost-a-evolucao-das-arvores-de-decisao/>

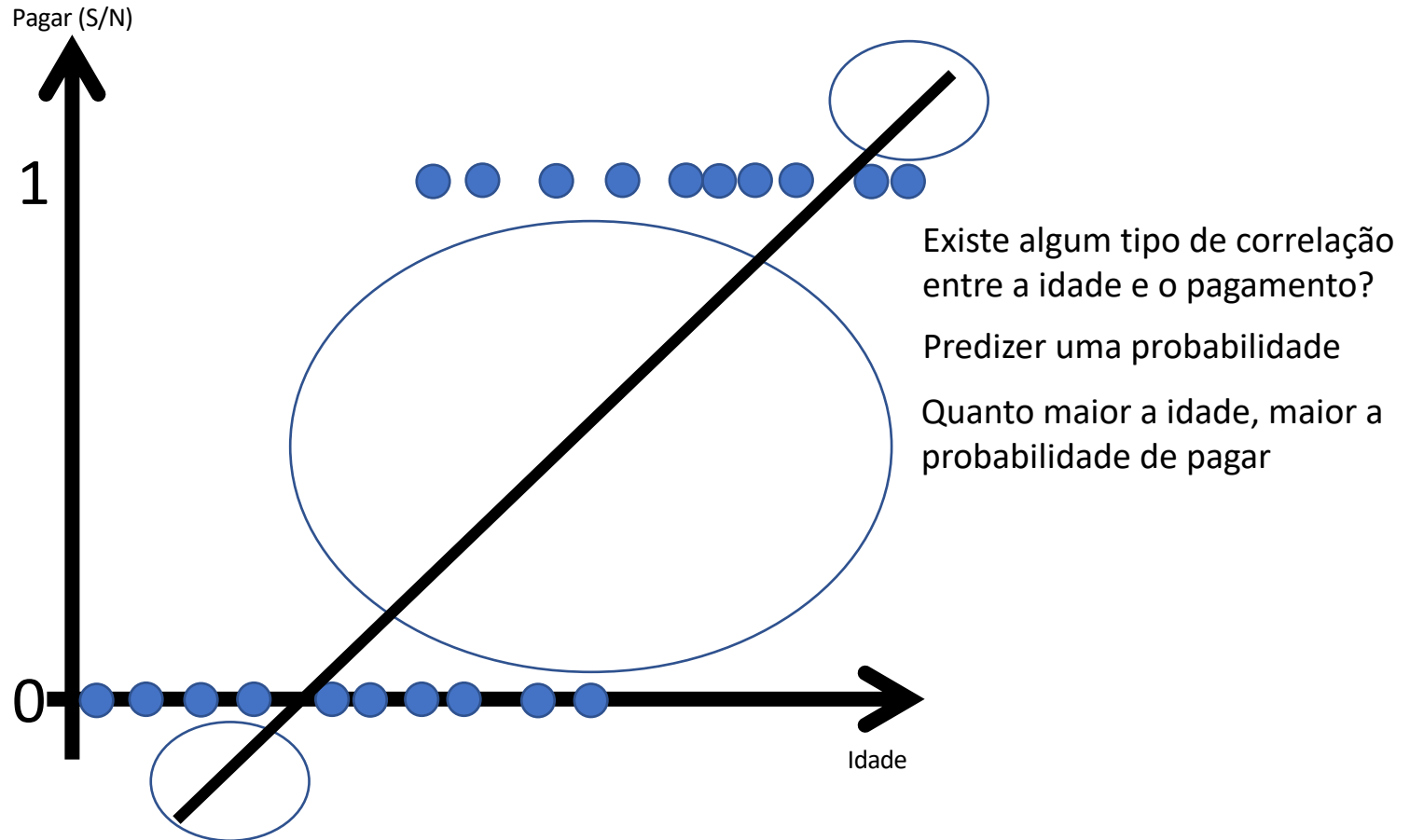
CLASSIFICAÇÃO

1. Base de dados do censo
2. Configurações do SageMaker
3. Linear learner (regressão logística)
4. Deploy do modelo
5. Exercício
 1. XGBoost
 2. Tuning do algoritmo

REGRESSÃO LOGÍSTICA



REGRESSÃO LOGÍSTICA

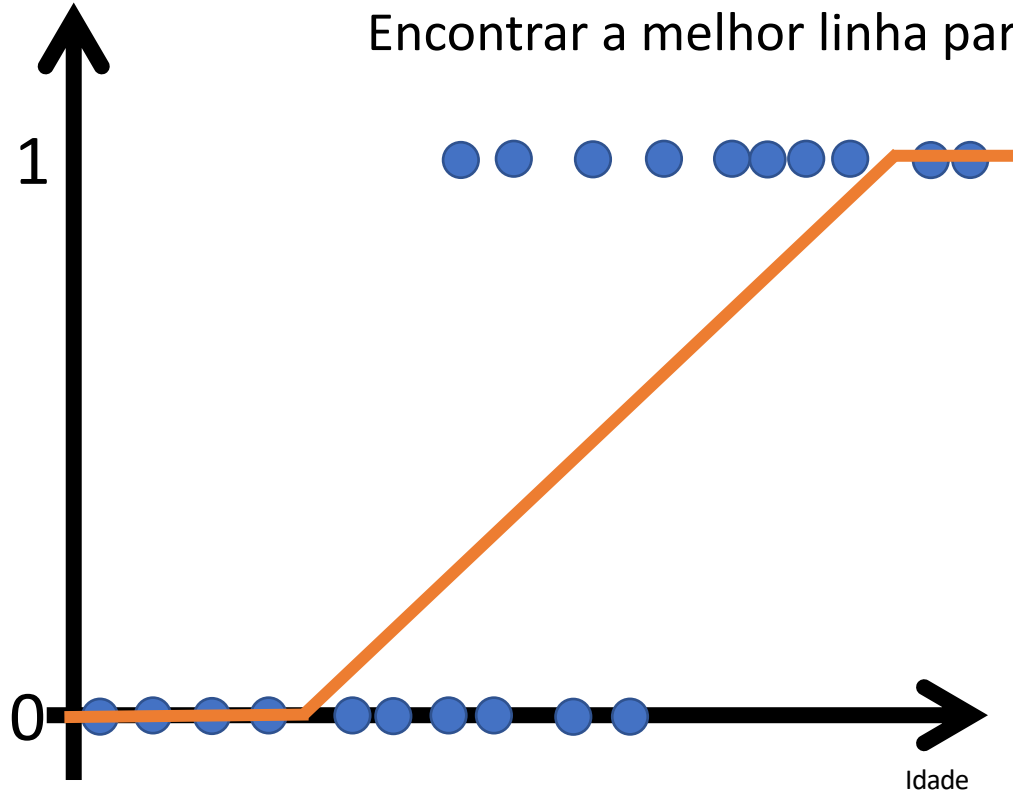


REGRESSÃO LOGÍSTICA

Pagar (S/N)

Função sigmoide

Encontrar a melhor linha para encaixar nos dados



REGRESSÃO LOGÍSTICA

$$y = b_0 + b_1 * x$$

x = idade

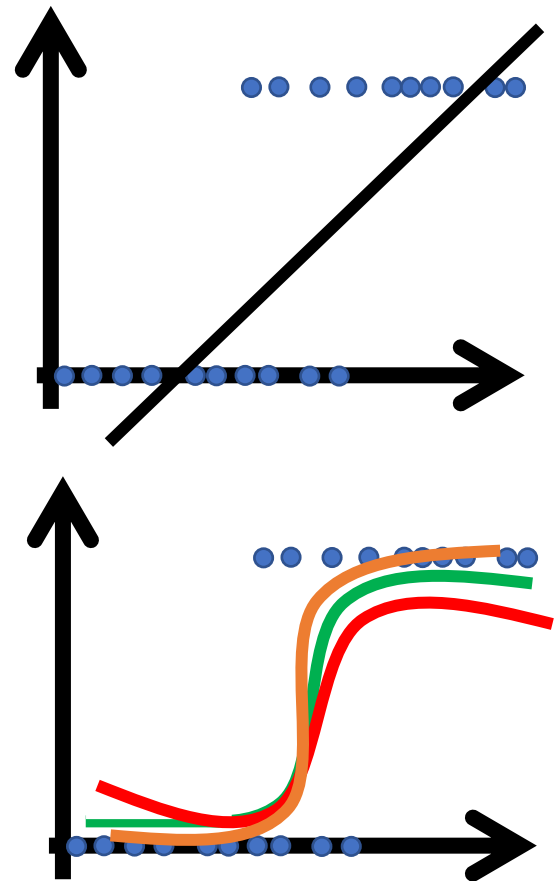
Equação da reta

$$p = \frac{1}{1 + e^{-y}}$$

Função sigmoide

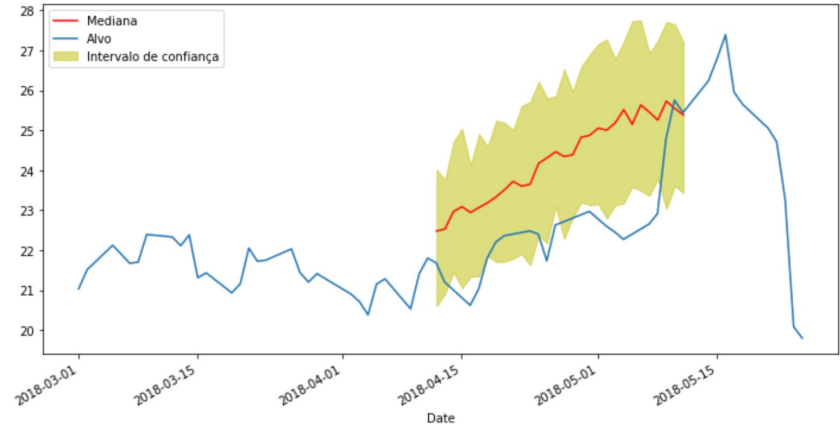
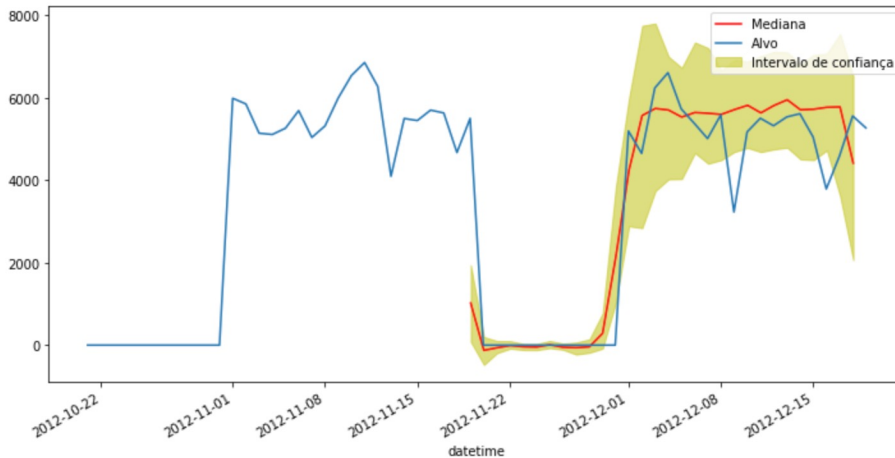
$$\log\left(\frac{p}{1-p}\right) = b_0 + b_1 * x$$

Transformação "logit"



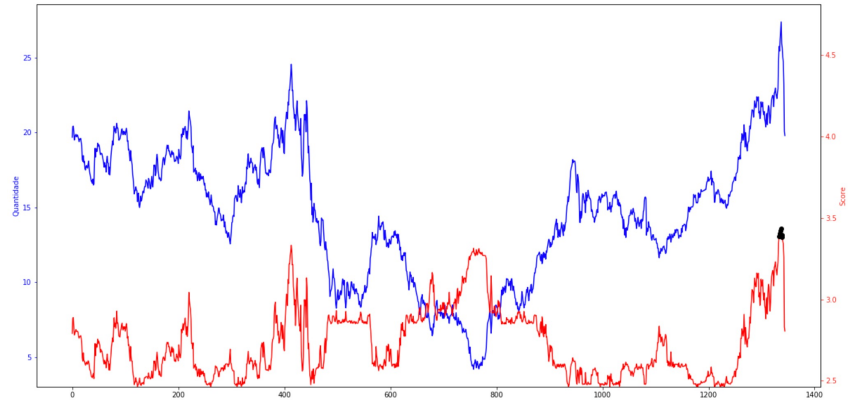
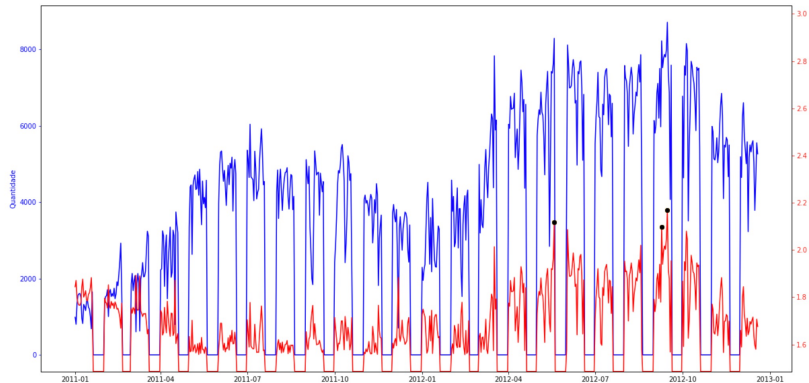
SÉRIES TEMPORAIS COM DEEPAR

1. Base de dados de aluguel de bicicletas
2. Base de dados de ações
3. Utiliza redes neurais recorrentes



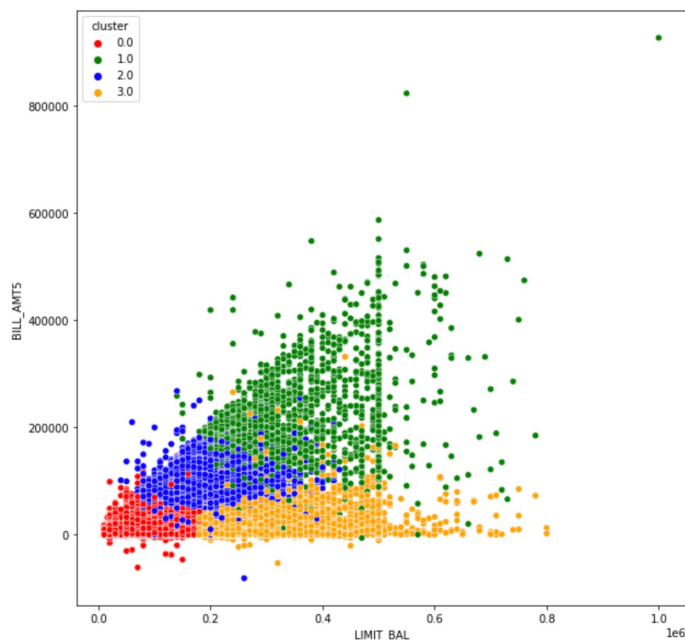
DETECÇÃO DE OUTLIERS COM RANDOM CUT

1. Base de dados de aluguel de bicicletas
2. Base de dados de ações



PCA (PRINCIPAL COMPONENT ANALYSIS) E AGRUPAMENTO COM K-MEANS

1. Base de dados do cartão de crédito: PCA + agrupamento
2. Base de dados do censo: PCA + classificação



REDES NEURAIS ARTIFICIAIS – IMAGE CLASSIFICATION

1. Aumento do limite de execução
2. Classificação de imagens com Caltech 256
3. Deploy e previsões
4. Base de dados MNIST – tipos de dados para redes neurais
5. Exercício

INTEGRAÇÃO TENSORFLOW E SAGEMAKER

1. Rede neural clássica com TensorFlow – base MNIST
2. Criação de script para executar o TensorFlow
3. Integração do TensorFlow com o SageMaker
4. Exercício

ENDPOINT EXTERNO

1. SageMaker SDK
2. Boto3 SDK
3. Base de dados das casas
4. Exercício – base MNIST

AUTOPILOT/EXPERIMENT (APRENDIZAGEM DE MÁQUINA AUTOMÁTICA)

1. Passos automáticos

- Pré-processamento dos dados
- Geração de algoritmos candidatos
- Feature engineering
- Tuning
- Geração de relatórios explicativos

2. Base de dados do censo

3. Base de dados das casas