

Bootcamp: Analista de Machine Learning

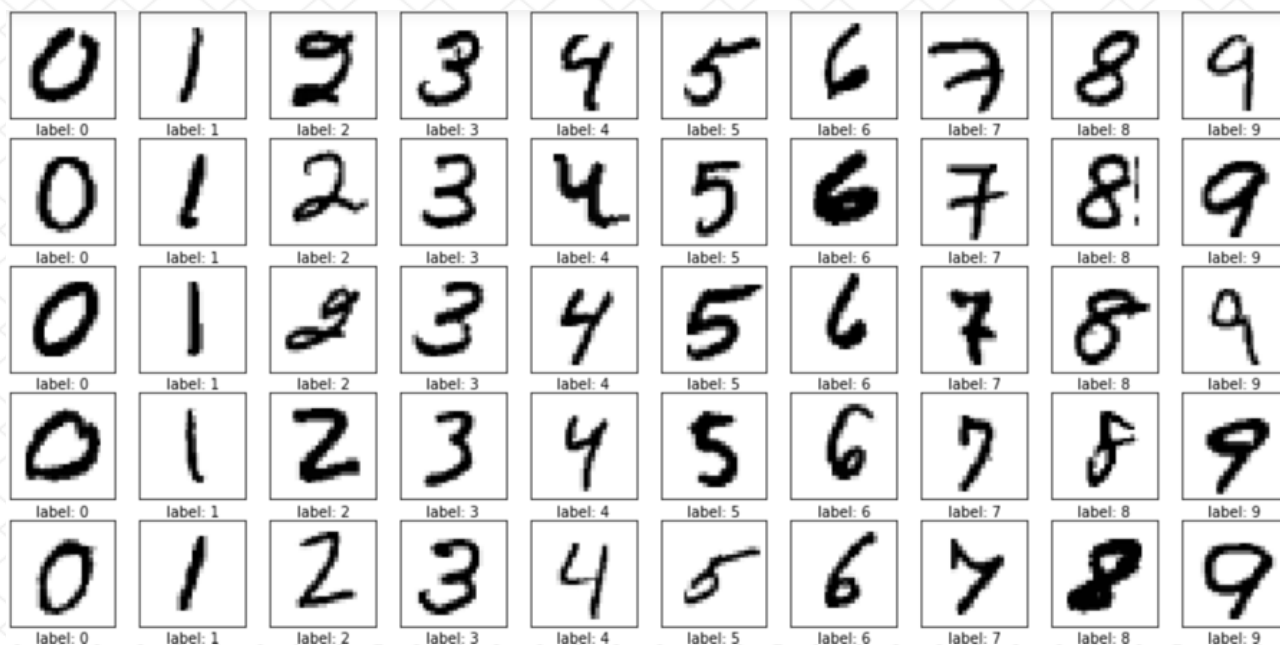
Desafio do módulo

Módulo 3	Seleção de Modelos de Aprendizado de Máquina
-----------------	---

Enunciado

A base de dados MNIST

Um dos mais clássicos conjuntos de dados para o treinamento e benchmark de algoritmos de classificação é o MNIST. Ele é composto por 70.000 imagens (60 mil de treino, 10 mil de teste) de números de 0 a 9 escritos à mão, já previamente classificados. Os números são sempre apresentados individualmente.



Suponha que você trabalha em uma empresa que faz a leitura automática (OCR) de formulários e cheques para instituições financeiras, e a você foi atribuído o trabalho de desenvolver um programa que faça a leitura de números escritos à mão nesses

documentos. Um tratamento prévio das imagens já entrega a você os números individualmente em imagens de resolução similar.

Cabe a você usar o aprendizado de máquina para ler cada imagem e dizer, com margem de segurança alta, que número foi escrito ali. Por falta de dados de teste, você usará o MNIST para treinar o seu algoritmo. Se seu algoritmo não tiver resultados excelentes, os números serão enviados para leitura e digitação manual, feita por operadores humanos, uma operação cara e pouco escalável.

Referências (acessos em 05/08/2020):

MNIST: <http://yann.lecun.com/exdb/mnist/>

Imagem: <https://web.archive.org/web/20180628145339/http://blog.welcomege.com/mnist-database/>

Respostas Finais

Os alunos deverão responder às seguintes questões objetivas: