



Machine Learning Crash Course

Traduzido por: Igor Freitas de Oliveira

Causa: Democratizar a Inteligência Artificial



Pré-requisitos e Pré-trabalho

O Curso Intensivo de Aprendizado de Máquina é ideal para você?

1. Eu tenho pouco ou nenhum conhecimento anterior em aprendizado de máquina.
2. Eu tenho algum conhecimento em aprendizado de máquina, mas gostaria de um entendimento mais atual e completo.
3. Eu sei muito bem o aprendizado de máquina, mas sei pouco ou nada sobre o TensorFlow.

Por favor leia completamente as seções seguintes de [Pré-requisitos](#) e [Pré-trabalhos](#) antes de começar o Curso Intensivo de Aprendizado de Máquina, para garantir que você está preparado para completar todos os módulos.

Pré-requisitos

O curso intensivo de aprendizado de máquina não presume ou requer nenhum conhecimento prévio em Aprendizado de Máquina. Porém, para entender os conceitos apresentados e completar os exercícios, recomendamos que os alunos conheçam aos seguintes pré-requisitos:

- **Domínio da Álgebra em nível introdutório.** Você deve se sentir confortável com variáveis, coeficientes, equações lineares, gráficos de funções e histogramas. (Familiaridade com conceitos mais avançados da matemática como logaritmos e derivadas é útil, mas não necessário).
- **Proficiência em programação básica e alguma experiência codificando em Python.** Exercícios de programação no curso intensivo de aprendizado de máquina são codificados em [Python](#) usando [TensorFlow](#). Não é necessária experiência anterior em TensorFlow, mas você deve sentir-se confortável lendo e escrevendo códigos Python que contém arquiteturas básicas da programação, como uma função definições/invocações, listas e dicionários, laços de repetição e expressões condicionais.

Começando com pandas

Os exercícios de programação no curso intensivo de aprendizado de máquina usam a biblioteca [pandas](#) para manipular os conjuntos de dados. Se você não é familiarizado com o pandas, recomendamos completar o tutorial [Introdução rápida ao pandas](#), que ilustra os principais recursos do pandas utilizados nos exercícios.

Fundamentos do TensorFlow de Baixo Nível

Os exercícios de programação no curso intensivo de aprendizado de máquina usa o TensorFlow de alto nível [tf.estimator API](#) para configurar os modelos. Se você está interessado em construir modelos do TensorFlow do zero, complete estes tutoriais:

- [TensorFlow Hello World](#): “Olá Mundo” codificado em baixo nível no TensorFlow.
- [TensorFlow Programming Concepts](#): Um passo a passo dos componentes fundamentais de uma aplicação do TensorFlow: tensores, operações, grafos e sessões.
- [Creating and Manipulating Tensors](#): Uma introdução rápida sobre tensores: A abstração central na programação do TensorFlow. Também fornece uma revisão sobre adição e multiplicação de matrizes em álgebra linear.

Conceitos e ferramentas chave

O curso intensivo de aprendizado de máquina discute e aplica os seguintes conceitos e ferramentas. Para mais informações, veja os recursos vinculados.

Matemática

Álgebra

- [Variáveis, coeficientes e funções](#)
- [Equações Lineares](#) como $y = b + x_1w_1 + x_2w_2$
- [Logaritmos](#) e equações logarítmicas como $y = \ln(1 + e^z)$
- [Função sigmoide](#)

Álgebra Linear

- [Tensor e grau de tensor](#)
- [Multiplicação de matriz](#)

Trigonometria

- [Tanh](#) (Discutido como uma [função de ativação](#); nenhum conhecimento prévio necessário)

Estatística

- [Média, mediana, outliers e desvio padrão](#)
- Habilidade para ler um [histograma](#)

Cálculo (Opcional, para tópicos avançados)

- Conceito de uma [derivada](#) (você não terá atualmente que calcular derivadas)
- [Gradiente](#) ou descida
- [Derivadas parciais](#) (que estão relacionadas exclusivamente aos gradientes)
- [Regra de cadeia](#) (Para entendimento completo do [algoritmo backpropagation](#) para treinamento das redes neurais)

Programação em Python

Python Básico

Os seguintes fundamentos do Python são abordados no [The Python Tutorial](#):

- [Definindo e chamando funções](#), usando parâmetros posicionais e [palavras-chaves](#)
- [Dicionários, listas, conjuntos](#) (criando, acessando e iterando)
- [Laços for](#), laços for com múltiplas variáveis de iteração (por exemplo, for a, b in [(1,2), (3,4)])
- [If/else blocos condicionais, expressões condicionais](#)
- [Formatação de String](#) (por exemplo, '%.2f' % 3.14)
- Variáveis, assinaturas, [tipos básicos de dados](#) (int, float, bool, str)
- A [declaração pass](#)

Python Intermediário

Os seguintes recursos mais avançados do Python também são abordados no [The Python Tutorial](#):

- [Compreensão de lista](#)
- [Funções lambda](#)

Bibliotecas Python de Terceiros

O curso intensivo de aprendizado de máquina codifica exemplos que usam os recursos das bibliotecas de terceiros. Familiaridade prévia com estas bibliotecas não é necessário; você pode procurar o que precisa saber quando precisar.

[Matplotlib](#) (para visualização de dados)

- [pyplot](#) módulo
- [cm](#) módulo
- [gridspec](#) módulo

[Seaborn](#) (para mapas de calor)

- [heatmap](#) função

[pandas](#) (para manipulação de dados)

- [DataFrame](#) classe

[NumPy](#) (para operações matemáticas em baixo nível)

- [linspace](#) função
- [random](#) função
- [array](#) função
- [arange](#) função

scikit-learn (para métricas de avaliação)

- [metrics](#) módulo

Terminal Bash / Cloud Console

Para rodar seus exercícios de programação na sua máquina local ou em um console na nuvem, você está confortável na linha de comando:

- [Bash Reference Manual](#)
- [Bash Cheatsheet](#)
- [Learn Shell](#)

Conceitos do Aprendizado de Máquina

Introdução ao Aprendizado de Máquina

Este módulo introduz o aprendizado de máquina

Tempo Estimado: 3 minutos

Objetivos do aprendizado:

- Reconhecer os benefícios práticos de dominar o aprendizado de máquina
- Entender a filosofia por trás do aprendizado de máquina

Exercício 1: Assista ao vídeo abaixo e faça as anotações descritas de acordo com o tópico objetivos do aprendizado.

[Introdução ao Aprendizado de Máquina](#)

Enquadramento

Este módulo investiga como enquadrar uma tarefa como um problema de aprendizado de máquina e abrange muitos dos termos básicos de vocabulário compartilhados por uma ampla gama de métodos de aprendizado de máquina (ML).

Tempo Estimado: 2 minutos

Objetivos do aprendizado:

- Atualizar os termos fundamentais do aprendizado de máquina.
- Explorar os vários usos do aprendizado de máquina

Exercício 2: Assista ao vídeo abaixo e faça as anotações descritas de acordo com o tópico objetivos do aprendizado.

[Enquadramento](#)

Enquadramento: Terminologia Chave do Aprendizado de Máquina

Tempo Estimado: 8 minutos

O que é aprendizado de máquina supervisionado? Concisamente é o seguinte:

- Sistemas de aprendizado de máquina aprende como combinar entrada para produzir previsões úteis sobre dados nunca vistos antes.

Vamos explorar a terminologia fundamental do aprendizado de máquina.

Rótulos

Um **rótulo** é uma coisa que estamos prevendo – A variável y na regressão linear simples. O rótulo pode ser o preço futuro do trigo, o tipo de animal mostrado em uma foto, o significado de um áudio ou praticamente qualquer coisa.

Características

Uma **característica** é uma variável de entrada – A variável x em uma regressão linear simples. Um projeto simples de aprendizado de máquina poderia usar uma única característica, enquanto um projeto de aprendizado de máquina mais sofisticado poderia usar milhões de características, especificadas como:

$$x_1, x_2, \dots x_N$$

No exemplo do detector de spam, as características podem incluir o seguinte:

- Palavras no texto do e-mail
- Endereço do remetente
- Hora do dia em que o e-mail foi enviado
- E-mail contendo a frase “Um truque estranho”

Amostras

Uma amostra é uma instância específica de dados, **x**. (Colocamos **x** em negrito para indicar que é um vetor.) Dividimos amostras em duas categorias:

- Amostras rotuladas
- Amostras não rotuladas

Uma **amostra rotulada** inclui característica(s) e um rótulo. Isto é:

Amostras rotuladas: {características, rótulo}: (x, y)

Use amostras rotuladas para treinar o modelo. Em nosso exemplo de detector de spam, as amostras rotuladas seriam e-mails individuais que usuários explicitamente marcaram como “spam” ou “não spam”.

Por exemplo, a tabela seguinte mostra 5 amostras rotuladas de um conjunto de dados contendo informações sobre os preços de imóveis na Califórnia:

housingMedianAge (feature)	totalRooms (feature)	totalBedrooms (feature)	medianHouseValue (label)
15	5612	1283	66900
19	7650	1901	80100
17	720	174	85700
14	1501	337	73400
20	1454	326	65500

Uma amostra rotulada contém características, mas não o rótulo. Isto é:

Amostras não rotuladas: {características, ?} : (x, ?)

Aqui estão 3 amostras não rotuladas do mesmo conjunto de dados de imóveis, que exclui medianHouseValue:

housingMedianAge (feature)	totalRooms (feature)	totalBedrooms (feature)
42	1686	361
34	1226	180
33	1077	271

Depois que treinamos nosso modelo com amostras rotuladas, usamos esse modelo para prever o rótulo em amostras não rotuladas. No detector de spam, amostras não rotuladas são novos e-mails que as pessoas ainda não rotularam.

Modelos

Um modelo representa a relação entre características e rótulos. Por exemplo, um modelo de detecção de spam pode associar determinados recursos a “spam”. Vamos destacar duas fases da vida de um modelo:

- **Treinando** significa criando ou **aprendendo** o modelo. Isto é, você mostra amostras rotuladas do modelo e permite que o modelo aprenda gradualmente as relações entre as características e o rótulo.
- **Inferência** significa aplicando o modelo treinado para amostras não rotuladas. Isto é, você usa o modelo treinado para fazer previsões úteis (y'). Por exemplo durante a inferência, você pode prever o `medianHouseValue` para novas amostras não rotuladas.

Regressão vs. Classificação

Um modelo de **regressão** prediz valores contínuos. Por exemplo, modelos de regressão fazem previsões que respondem questões como as seguintes:

- Qual é o valor de uma casa na Califórnia?
- Qual é a probabilidade de um usuário clicar neste anúncio?

Um modelo de classificação prediz valores discretos. Por exemplo, modelos de classificação fazem previsões que respondem questões como as seguintes:

- Dado uma mensagem de e-mail é spam ou não spam?
- Esta é uma imagem de um cachorro, um gato ou um hamster?

Recursos e Informações adicionais

Fontes

Tempo Estimado:

Hexadecimal: #0288d1

RGB: 2 136 209

Link:

Hexadecimal: #009BE5

RGB: 0 155 229

Objetivos de Aprendizado:

Hexadecimal: #00897b

RGB: 0 137 123

Exemplos:

Hexadecimal: #37474f

RGB: 55 71 79

Fonte: Roboto Mono

Títulos:

Hexadecimal: #212121

RGB: 33 33 33

Termos Chave:

Hexadecimal: #9c27b0

RGB: 33 33 33