



## Sobre o curso

---

Fabrcio Olivetti de Franca

Universidade Federal do ABC

1. Sobre mim
2. Sobre a Disciplina

## Sobre mim

---

Formado em Engenharia Elétrica com ênfase em Computação pela Universidade Católica de Santos.

Mestrado e Doutorado no mesmo curso na Universidade Estadual de Campinas.

Áreas de pesquisa:

- Co-agrupamento de Dados
- Regressão Simbólica
- Algoritmos Evolutivos
- Heurísticas para Otimização
- Aprendizado de Máquina

## **Sobre a Disciplina**

---

O avanço da tecnologia nos permitiu obter dados em massa de diversas fontes:

- Transações Bancárias
- Dados de medidas de sensores
- Experimentos genéticos
- Uso de Redes Sociais
- Construção de conteúdo colaborativo

Desses dados surge a necessidade de extrair informações úteis:

- Possíveis dívidas bancárias
- Genes relacionados a doenças
- Previsão de desastres naturais



- Muitos algoritmos consolidados
- Foco em bases de dados pequenas e bem estruturadas
- Bases de dados pequenas em apenas uma máquina!

- Ser capaz de trabalhar como Cientista de Dados.
- Entender os três tipos de Aprendizado de Máquina: Supervisionado, Não-Supervisionado e por Reforço.
- Aprender algoritmos básicos e avançados.

<http://folivetti.github.io/teaching/2018-summer-teaching-1>

- Aulas expositivas
- Atividades de Programação (baseados em material fornecido pela Intel)

- Distribuição Intel da linguagem Python (via Anaconda).
- Binder para algumas demonstrações sem instalação.

No site.

Utilizaremos o Jupyter Notebook com a linguagem Python e as bibliotecas numéricas otimizadas pela Intel.

Siga passo a passo o tutorial de instalação:

<https://software.intel.com/en-us/articles/using-intel-distribution-for-python-with-anaconda>

Livros sobre Mineração de Dados:

Mining Massive Datasets

Introduction to Data Mining – P-N. Tan, M. Steinbach, V.Kumar –  
Addison Wesley 2005.

Data mining : practical machine learning tools and techniques / 3. ed.



Livros sobre Python:

LUTZ, Mark; ASCHER, David; ALYH69. Aprendendo python. 2. ed.  
Porto Alegre, RS: Bookman, 2007. 566 p. ISBN 857780013x.

As avaliações consistirão de:

- Entrega das Atividades (1 pt)
- 02 Provas (6 pts)
- Um projeto final (3 pts)

As atividades são de aplicação prática em cada tipo de aprendizado estudado na disciplina.

Utilizaremos a biblioteca *scikit-learn* para aplicação dos algoritmos.

A implementação dos algoritmos fica como tarefa opcional e incentivada (afinal vocês adoram programar).

Para Aprendizado Supervisionado e Não-Supervisionado:

- Predição de venda de imóveis.
- Reconhecimento de atividade humana em smartphones.
- Análise de perda de funcionários em uma empresa.
- Predição de cancelamento de assinaturas de empresa telefônica.
- Predição de vendas de liquidação.

As atividades se encontram em:

<https://github.com/folivetti/AprendizadoMaquina>

E podem ser feitas online pelo endereço:

<https://mybinder.org/v2/gh/folivetti/AprendizadoMaquina/master>

(nota, é necessário Salvar o notebook e clicar em File → Download as Notebook para gerar o arquivo para entrega)

Cada aluno deve criar uma conta no github e criar um repositório UFABC-ML-2018

Nesse repositório, insira as atividades dentro do prazo (dia da prova correspondente ao assunto).

Na próxima semana enviar o link de seus repositórios para folivetti @ ufabc.edu.br

No dia da entrega as atividades serão pontuadas. Não faça "commit" após o prazo.

Para Aprendizado por Reforço utilizaremos a interface Retro Learning Environment para aprender a jogar o jogo Super Mario World de SNES.

Esse será o projeto final que será graduado em estilo de competição. Detalhes serão dados após a segunda prova.



O conceito final será calculado da seguinte forma:

$$A \geq 8pts$$

$$7pts \leq B < 8pts$$

$$6pts \leq C < 7pts$$

$$5pts \leq D < 6pts$$

$$F < 5pts$$

A prova de recuperação substituirá a menor nota dentre as 3 provas.

Caso esse critério não seja suficiente para subir o conceito, será acrescentada uma atividade extra envolvendo participação de uma competição no Kaggle, a ser anunciada na data da recuperação.