

Plano de Ensino: Aprendizado de Máquinas

Saul Leite
Centro de Matemática, Computação e Cognição
Universidade Federal do ABC

20 setembro, 2020

Detalhes:

- Sigla: MCZA002-17
- T-P-I: 4-0-4
- Carga horária: 48 horas
- Recomendações: Inteligência Artificial

Recomendações para essa turma:

- Processamento da Informação.
- Introdução a Probabilidade e à Estatística.
- Funções de uma Variável (FUV).

Sobre o Professor

- Nome: Saul Leite
- Email: saul.leite@ufabc.edu.br
- Site: professor.ufabc.edu.br/~saul.leite

Objetivos da Disciplina

- (i) Assimilar conceitos para a utilização de algoritmos e técnicas de aprendizados de máquina;
- (ii) Apresentar/experimentar uso desses algoritmos e técnicas em diferentes domínios de aplicação, incluindo a resolução de problemas de natureza científica e/ou tecnológica, enfatizando o estudo comparativo desses algoritmos.

Conteúdo Programático:

Introdução. Tipos de aprendizado. Paradigmas de aprendizado. Avaliação experimental de algoritmos de Aprendizado de Máquina. Alguns algoritmos de Aprendizado de Máquina.

Programação das Aulas

Semana 01:

- Detalhes sobre a disciplina
- Introdução ao R

Semana 02

- Introdução ao Aprendizado de Máquinas
- Introdução ao Aprendizado Supervisionado, algoritmo k NN

Semana 03

- Viés-Variância, Validação de Modelos
- Regressão Linear

Semana 04

- Seleção de Características
- Regressão Regularizada, Ridge, Lasso, Elastic-Net

Semana 05

- Regressão Não-Linear, Funções Base, Regressão Local
- Kernels positivos definidos, Regressão Ridge com Kernel

Semana 06

- Análise Discriminante
- Classificadores Lineares, Perceptron, métodos on-line e em batch

Semana 07

- Máquinas de Vetores Suporte
- Medidas de Avaliação, Múltiplas Classes

Semana 08

- Árvores para Regressão e Classificação
- Bagging, Boosting e Random Forests

Semana 09

- Redes Neurais Artificiais *feedforward*
- Redes Neurais Convolucionais

Semana 10

- Agrupamento, K-médias, Hierárquico
- Análise dos componentes Principais (PCA)

Semana 11

- Aprendizado por Reforço, Definições e Método Monte-Carlo
- Algoritmos Q-Learning e Sarsa

Semana 12

- Apresentação de Projetos
- Apresentação de Projetos

Semana 13

- 15/12 - *REC*

Outras Informações:

Materiais necessários para disciplina

Um computador com conexão com a Internet, capaz de rodar RStudio e R, microfone para fazer vídeos.

- RStudio é uma ferramenta opensource que pode ser baixada gratuitamente no link abaixo:
<https://rstudio.com/>
- R pode ser baixado no link:
<https://www.r-project.org/>
- Alternativamente, você pode fazer download da pacote Anaconda, que possui opção de instalar RStudio, R (e Python, caso deseje usar na disciplina). O link está abaixo:
<https://www.anaconda.com/>

Referências Bibliográficas

- [1] J. Gareth, D. Witten, T. Hastie, R. Tibshirani, “An Introduction to Statistical Learning: with Applications in R”, Springer 2013. (Disponível em: <http://faculty.marshall.usc.edu/gareth-james/ISL/>)
- [2] M. Mohri, A. Rostamizadeh, A. Talwalkar, “Foundations of Machine Learning”, MIT Press, 2012. (Disponível em: <https://cs.nyu.edu/~mohri/mlbook/>)
- [3] T. Hastie, R. Tibshirani, J. Friedman, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction”, Second Edition, Springer, 2009. (Disponível em <https://web.stanford.edu/~hastie/ElemStatLearn/>)

Site do Curso

Google Classroom

O gerenciamento da disciplina será feito através do Google Classroom.

- Vídeo Aulas
- Avisos
- Listas

Um convite será enviado para a conta de todos os alunos matriculados. Você deve lembrar de fazer *login* usando seu email **institucional** da UFABC.

Horário de Atendimento

Discord

O atendimento para esta disciplina ocorrerá no canal do Discord abaixo:

<https://discord.gg/9MYzReC>

Os horários de atendimento são:

- Terça-Feira das 19:00 às 20:00
- Sexta-Feira das 21:00 às 22:00

Forma de Avaliação

A avaliação da disciplina será dividida nas seguintes **categorias**:

- Listas Semanais - 30%
- Vídeo Soluções - 20%
- Projeto Final - 30%
- Avaliação por Pares - 20%

Ao lado de cada avaliação estão os **pesos** correspondentes a cada categoria.

Conceitos usados para avaliação

Todas as atividades serão avaliadas de acordo com os seguintes conceitos:

- A: **Trabalho excepcional;**
- B: **Bom desempenho no trabalho;**
- C: **Desempenho adequado;**
- D: **Desempenho mínimo;**
- F: **Desempenho insuficiente;**

O resultado de cada **categoria** será calculado convertendo os conceitos para a escala numérica de 4 a 0, representando conceitos de A à F e calculando uma média aritmética.

Nota final

A nota final N será computada como a média dos resultados de cada **categoria**, ponderada pelos pesos correspondentes. O valor final será convertido para conceitos de acordo com a seguinte tabela:

- A: se $N \in [3.5, 4]$
- B: se $N \in [2.8, 3.5)$
- C: se $N \in [1.9, 2.8)$
- D: se $N \in [1.0, 1.9)$
- F: se $N < 1.0$

Detalhes sobre Listas Semanais

- Feitas usando o **GitHub Classroom**.
 - Soluções devem ser feitas usando **RMarkdown** e compilado em **PDF**.
- Serão feitas em **duplas** escolhida através de um **sorteio**.
 - Os integrantes da dupla devem conhecer e participar da resolução de cada exercício, mas a divisão de tarefas fica por conta dos participantes.
- Alguns exercícios serão listados como *opcionais*. Estes não precisam ser entregues junto com a lista, mas os alunos podem entregá-los se desejarem.
- Como o número de alunos pode ser ímpar, um aluno poderá ser sorteado para entregar a lista sozinho. Cada aluno somente ficará fora de uma dupla uma vez.

Detalhes sobre as Vídeo Soluções

- Cada aluno deverá fazer **dois** vídeos apresentado sua solução de um exercício da lista de até no **máximo 5 minutos**.
 - Os vídeos serão disponibilizado para todos os alunos da turma em canal de **Youtube não listado** ou na página da disciplina no **Google Classroom**.
 - Um vídeo deverá ser sobre exercícios das listas das semanas de 02 à 06;
 - Outro vídeo deverá ser sobre exercícios das listas das semana de 07 à 11.
- Escolha do exercício para resolver:
 - Cada exercício poderá ser resolvido por até dois alunos (**não** da mesma dupla).
 - Todos estão livres para se candidatar a apresentar suas soluções - os primeiros dois candidatos a entrarem em contato serão escolhidos para apresentar.
 - Se não houverem candidatos, um sorteio será realizado pelo professor e os alunos sorteados terão 72 horas para fazer o vídeo e entregar suas soluções.
 - Alunos sorteados para resolução individual da lista não serão sorteados.
 - Os exercícios *opcionais* não serão sorteados, mas os alunos que desejarem poderão apresentá-los para cumprir essa atividade.

Detalhes sobre Projetos Finais

- Projetos finais serão realizados utilizando o **GitHub Classroom** com grupos de até 5 integrantes.
- Os grupos serão formados pelos próprios alunos.
- Consistirá na aplicação e comparação de métodos de Machine Learning em bancos de dados públicos. Materiais para entregar:

- Um relatório usando o RMarkdown contendo o código desenvolvido;
 - Um vídeo de apresentação do projeto de até 15 minutos. O vídeo deve ter a participação de todos os integrantes do grupo e será disponibilizado para todos os alunos da disciplina pelo **Youtube em canal não listado** ou no **Google Classroom**.
- Mais detalhes serão apresentados no decorrer da disciplina.

Detalhes sobre a Avaliação por Pares

- Após a entrega de cada lista e do projeto final, todos alunos participantes devem avaliar a participação dos seus colegas de grupo com conceitos de A à F.
- As avaliações serão feitas através de formulários disponibilizados no Google Classroom.

ATENÇÃO: A falha de preenchimento dos formulários **no prazo estipulado** implicará em anulação da nota na atividade correspondente.

Honestidade Acadêmica

- As listas e projetos devem ser realizadas somente pelos integrantes do grupo.
- Livros e a Internet poderão ser consultados para todas as atividades, mas devem ser devidamente citados e o texto o final e os códigos devem ser de **sua autoria**.
- Integrantes dos grupos não devem comunicar com outros alunos da turma, exceto através de canais públicos de discussão da disciplina no Discord.
- **Detector de plágio** para texto e código será utilizado para todas as soluções.

ATENÇÃO: Todas as providências cabíveis serão tomadas em caso de detecção de plágio para **todos** os envolvidos, incluindo:

- Descarte dos conceitos atribuídos a TODAS as tarefas avaliativas regulares de TODOS os envolvidos, causando assim suas reprovações automáticas com conceito F.
- Possível denúncia à Comissão de Transgressões Disciplinares Discentes da Graduação, a qual decidirá sobre a punição adequada à violação que pode resultar em advertência, suspensão ou desligamento, de acordo com os artigos 78-82 do Regimento Geral da UFABC.
- Possível denúncia apresentada à Comissão de Ética da UFABC, de acordo com o artigo 25 do Código de Ética da UFABC.

Para maiores detalhes veja <http://professor.ufabc.edu.br/~e.francesquini/codigodehonra/>.

Recuperação:

Alunos com conceitos finais D ou F poderão fazer a avaliação REC.

- Ocorrerá entre o período de 14 à 19 de Dezembro.
- Consiste em uma lista de exercícios.
- Deverá ser entregue a solução usando RMarkdown junto com a apresentação da solução via **vídeo**.
- A prova terá duração de 72 horas.

A nota final após REC será computada de acordo com a seguinte tabela:

Antes da Rec	Rec	Final
D	A	C
D	B	C
D	C	D
D	D	D
D	F	F
F	A	C
F	B	D
F	C	F
F	D	F
F	F	F