

Introdução a Aprendizagem de máquina para Bioinformática

Aula 0

Msc. Amanda Araújo Serrão de Andrade

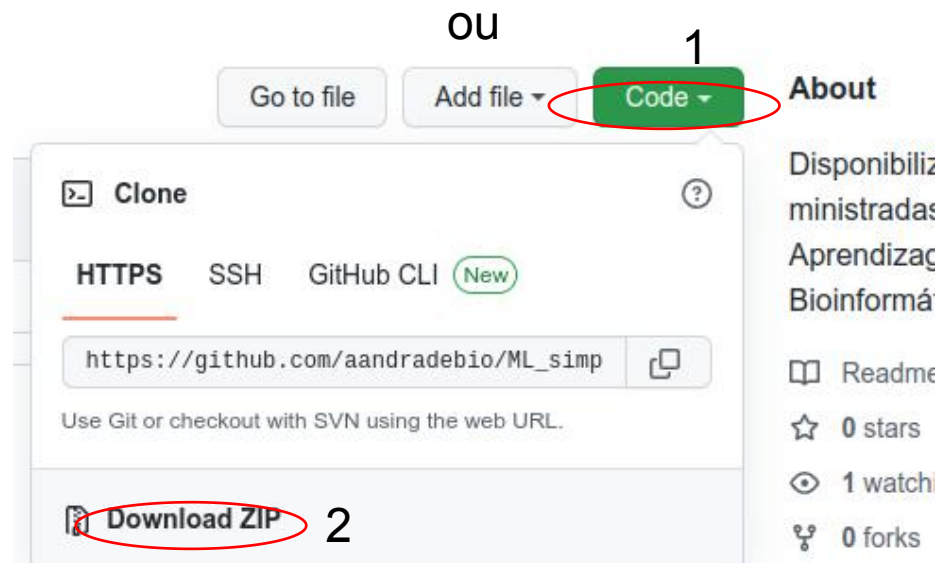
Instalação do R e do RStudio

- Link para tutoriais:
 - 1) <https://www.dataquest.io/blog/tutorial-getting-started-with-r-and-rstudio/>
 - 2) <https://rstudio-education.github.io/hopr/starting.html>

Repositório do minicurso

- https://github.com/aandradebio/ML_simposioPGGEN
- Foi criado um repositório que contem todas as aulas práticas e conteúdos adicionais relacionado a este minicurso.
- Sugiro o download antes do inicio das aulas:

```
$ wget https://github.com/aandradebio/ML_simposioPGGEN/archive/refs/heads/main.zip
```



Instalação das dependências

- Dependências através do script `install_dependencies.R` depositado no GitHub ou através do comando abaixo:
- `install.packages(c("seqinrR","kmer", "Caret", "MLeval", "ggplot2", "ggtree", "dplyr", "ape", "tidyverse","e1071", "randomForest", "ranger", "tidyr", "adabag", "extraTrees","ISLR","caretEnsemble"))`

Conteúdo das próximas aulas

Aula 1 (26/09)
<ul style="list-style-type: none">• Obtenção dos conjuntos de dados• Representações numéricas de sequências biológicas• Filtragem e pré-processamento dos conjuntos de dados• Desbalanceamento
Aula 2 (27/09)
<ul style="list-style-type: none">• Desbalanceamento• Validação cruzada• Classificação• Clusterização e Regressão
Aula 3 (28/09)
<ul style="list-style-type: none">• Principais métricas de avaliação• Resultados• Comparação de performance

Referências

- <https://machinelearningmastery.com/machine-learning-in-r-step-by-step/>
- <https://topepo.github.io/caret/>
- https://github.com/aandradebio/ML_simposioPGGEN
- <https://www.dataquest.io/blog/tutorial-getting-started-with-r-and-rstudio/>



Dúvidas?

aandradebio@gmail.com

- As dúvidas mais específicas devem vir acompanhadas do código e de um conjunto de dados de teste para agilizar a resposta



Obrigada!!

aandradebio@gmail.com

atr@Incc.br