

# Aula 1 - Introdução à ciência de dados

Thiago S. F .Silva

# Apresentações

Eu: Thiago Silva

- ▶ Biologia (UFRN, 2002)
- ▶ Sensoriamento Remoto (INPE, 2004)
- ▶ Geografia Física (UVic, 2009)
- ▶ Funcionamento e dinâmica de ecossistemas
- ▶ Efeitos das mudanças globais sobre a biota
- ▶ Interface entre Ecologia, Computação e Geociências
- ▶ “Ecologia Digital”

Vocês?

# O que esse curso é sobre?

- ▶ Análise **preditiva** de dados
- ▶ Introdução ao aprendizado de máquina (*machine learning*)
- ▶ Conversa sobre princípios e práticas científicas
- ▶ Ponto de partida para o seu próprio aprendizado

# O que esse curso não é sobre?

- ▶ Inferência estatística (mas vamos tocar nela várias vezes)
- ▶ “IA”

*Machine Learning?*



# Artificial Intelligence

Automated Programming   Knowledge Representation   Expert Systems

Planning and Scheduling

Intelligent Robotics

Visual Perception

Speech Recognition

## Machine Learning

K-Means Clustering

Principal Component Analysis (PCA)

Automatic Reasoning

Problem Solving & Search Strategies

Decision Trees

Random Forest

Natural Language Processing (NLP)

K-Nearest Neighbors (KNN)

Ensemble Methods

Support Vector Machines (SVM)

Naive Bayes Classification

Linear / Logistic Regression

Radial Basis Function Networks

Recurrent Neural Networks (RNN)

Autoencoders

Anomaly Detection

Self-Organizing Maps

Hopfield Networks

Reinforcement Learning

Multilayer Perceptrons (MLP)

Modular Neural Networks

## Deep Learning

Boltzmann Machines

Adaptive Resonance Theory (ART)

Convolutional Neural Networks (CNN)

Long Short-Term Memory Networks (LSTM)

Recurrent Neural Networks (RNN)

Deep Reinforcement Learning

Generative Adversarial Networks (GAN)

Transformer Models (such as BERT, GPT)

Deep Autoencoders

Deep Belief Networks (DBN)

# Estatística?

Matemática vs. Estatística?

O que é Estatística?

Por que usamos Estatística?



# Estatística

- ▶ Exploração
- ▶ Confirmação
- ▶ Predição

# Estatística

- ▶ Exploração: sem expectativas prévias, perguntas abertas.
- ▶ Confirmação: hipóteses pré-estabelecidas
- ▶ Predição: foco em gerar novos dados

## Exemplo:

**Exp:** Quão intenso é o efeito da redução de disponibilidade hídrica sobre o crescimento da espécie de planta *Plantus plantus*?

**Conf:** A redução em 50% na disponibilidade hídrica esperada para a década de 2100 resultará em redução de pelo menos 30% no crescimento de *Plantus plantus*.

**Pred:** Qual o crescimento esperado para *Plantus plantus* em 2100 sob os cenários A, B e C de mudanças climáticas?

# Machine Learning

- ▶ Utiliza análise exploratória para entendimento dos *dados*
- ▶ Não se preocupa com confirmação
- ▶ Especializado em gerar as melhores *predições* possíveis.

# Machine learning é estatística?



