HPE DSI 311
Introduction to Machine Learning

Spring 2023

Instructor: Ioannis Konstantinidis

Overview

- What topics will we focus on in this introductory class?
- DETOUR: what topics are beyond the introductory level?
- Brief historical context

Topics we will focus on in this introductory class

Data: Structured, Toy Examples

- Flat, static CSV files
- Tidy (normalized, tabular)
- Clean; no parts
  - incomplete,
  - incorrect,
  - inaccurate, or
  - irrelevant

Models: training and evaluation

- Regularization
- Cross-validation
- Objective functions
- Scoring methods

Models: mostly for supervised learning

Models: also a few unsupervised learning ones

K-means

Hierarchical clustering

Model extensions
FORESTS
MACHINES
PIPELINES
KERNELS

Deep Learning: A small taste of

Software Environment:

Hewlett Packard Enterprise
Data Science Institute
UNIVERSITY OF HOUSTON

That's plenty for five weeks!

Other/ Advanced ML topics

A parallel universe of ML software

But that's a whole 'nother course …

Ingredients for success

Hewlett Packard Enterprise
Data Science Institute
UNIVERSITY OF HOUSTON

DATA

Data Format

Data Source	Structured	Unstructured
Internal	**Human-Generated** • Survey ratings • Aptitude testing **Machine-Generated** • Web metrics from Web logs • Product purchase from sales Records • Process control measures	**Human-Generated** • Emails, letters, text messages • Audio transcripts • Customer comments • Voicemails • Corporate video/communications • Pictures, illustrations • Employee reviews
External	**Human-Generated** • Number of Retweets, Facebook likes, Google Plus +1s • Ratings on Yelp • Patient ratings ratings **Machine-Generated** • GPS for tweets • Time of tweet/updates/postings	**Human-Generated** • Content of social media updates • Comments in online forums • Comments on Yelp • Video reviews • Pinterest images • Surveillance video

Copyright 2014 Business Over Broadway

DATA

Data
Velocity
Real Time
near Real Time
periodic
batch
table
MB
GB
TB
PB
Data
Volume
data base
photo
Web
Audio
Social
video
unstructured
Mobile
Data
Variety

Curation / Preprocessing

These are topics that this class will not cover!

HPE DSI offers “262 - Principles of Data Management” as an introduction

Choose the right model … from many options

MODELS

A mostly complete chart of
Neural Networks
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probabilistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)

Convolutional Neural Networks

Convolutional NN is a multi-layered neural network designed to extract increasingly complex features of the data at each layer.

https://www.guru99.com/deep-learning-tutorial.html

Attention/ Transformers

MODELS

These are topics that this class will not cover!

HPE DSI offers “312 - Introduction to Deep Learning” as an introduction

SOFTWARE TOOLS

Parallelization of Model Implementation

- across multiple cores on a single machine
- using GPUs (e.g. CUDA/OpenCL)
- across multiple nodes in a cluster (e.g. MLlib on Spark, Mahout on Hadoop/Mapreduce)

Carya @ HPEDSI

Model Deployment in Production (MLOPS)

Servers
Deploy in months

Virtual Machines
Deploy in minutes

Containers
Deploy in seconds

Serverless
Deploys in milliseconds

UI/UX

- Dashboards (Plotly/Dash/Tableau)
- Javascript dataviz (React + D3)
- Python Web apps (Django/Flask)

These are topics that this class will not cover!

HPE DSI offers “251 - Data Visualization” as an introduction

END
DETOUR

Brief historical perspective for context

What's in a name?

That which we call a rose
by any other name
would smell as sweet

What's in a name?

That which we call ~~*a rose*~~ *ML*
by any other name
would smell as sweet:

- AI
- Data Science
- Statistics
- Data Analytics
- Approximation Theory (Math)
- Data Mining

AI/ML is not new

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

1. *The Imitation Game.* AKA the Turing Test

I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to

AI/ML is not new

Connectionist AI:
Neural networks,
data corpora

BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 5, 1943

A LOGICAL CALCULUS OF THE
IDEAS IMMANENT IN NERVOUS ACTIVITY

WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE,
DEPARTMENT OF PSYCHIATRY AT THE ILLINOIS NEUROPSYCHIATRIC INSTITUTE,
AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propositional logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expression satisfying certain conditions, one can find a net behaving in the fashion it describes. It is shown that many particular choices among possible neurophysiological assumptions are equivalent, in the sense that for every net behaving under one assumption, there exists another net which behaves under the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

https://neurovenge.antonomase.fr/

AI/ML is not new, but it comes in two flavors

Symbolic AI:
Expert Systems,
Data Ontologies

https://neurovenge.antonomase.fr/

AI/ML is not new, but it comes in two flavors

Symbolic AI:
Expert Systems,
Data Ontologies

A PROPOSAL FOR THE
DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE

"The study is to proceed on the basis of the conjecture
that every aspect of learning or
any other feature of intelligence
can in principle be so **precisely described** that
a machine can be made to simulate it."

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I. B. M. Corporation
C. E. Shannon, Bell Telephone Laboratories

Early chatbot example of symbolic AI

Symbolic AI:
Expert Systems,
Data Ontologies

Joseph Weizenbaum's program ELIZA, published in 1966, simulated conversation by using a "pattern matching" and substitution methodology. **Directives on how to interact** were provided by "scripts", which allowed ELIZA to process user inputs and engage in discourse following the **rules and directions of the script**, using pre-prepared or pre-programmed responses.

For example:

EXPERT KNOWLEDGE: the meaning of the word ‘FAMILY’ is broader than the meaning of the word ‘MOTHER’

RULE: Respond to any input that contains the word 'MOTHER’ with 'TELL ME MORE ABOUT YOUR FAMILY’

https://sites.google.com/view/elizagen-org/the-original-eliza

ML skews to a different flavor of AI

Artificial Intelligence (AI)

Symbolic AI:
Expert Systems,
Data Ontologies

Machine Learning (ML):
refers to (mostly connectionist) AI techniques that do not rely on explicit rules based on expert knowledge, but on **patterns inferred from data examples** instead

ML skews to the patterns flavor of AI

https://neurovenge.antonomase.fr/

Connectionist AI:
Neural networks,
big data corpora

Modern chatbot example

Collect as much written text as possible
For any given input (consecutive words),
find the closest match among the collected data, and
respond with the word that follows in the matched text

This is essentially a version of the “nearest neighbor” algorithm, and is used by state of the art NLP systems (e.g., chatGPT)

CAUTION: this is an (extremely oversimplified) example, not an actual blueprint for how all NLP systems work

ML skews to the "patterns" flavor of AI

Artificial Intelligence (AI)

Symbolic AI:
Refers to techniques that rely on **explicit rules based on expert knowledge** about data ontologies

Machine Learning (ML): refers to (mostly connectionist) techniques that rely on **patterns inferred from data examples**

ML skews to the "patterns" flavor of AI

Start with **THEORY** to **construct** MODEL

Artificial Intelligence (AI)

Symbolic AI:
Refers to techniques that rely on **explicit rules based on expert knowledge** about data ontologies

Machine Learning (ML): refers to (mostly connectionist) techniques that rely on **patterns inferred from data examples**

Start with **DATA** to **fit** MODEL

Big Data + Computing at Scale = Deep ML

Artificial Intelligence (AI)

Symbolic AI:
Refers to techniques that rely on **explicit rules based on expert knowledge** about data ontologies

Machine Learning (ML):

Deep Learning relies on fitting LARGE models (million+ parameters)

Train:
Dog
Cat
Errors
Dog
Cat
Raccoon
Deploy:
Dog

Neural Networks (e.g., Perceptron, 1950s)

Train:
Error Backpropagation (1980s)
Dog
Cat
Neural Networks (e.g., Perceptron, 1950s)
Dog
Cat
Raccoon
Deploy:
Dog

The missing ingredients?

Big Data + Computing at Scale

The technology (GPUs) and dataset sizes (Big Data) needed for practical implementations of Deep Learning algorithms did not mature until the last decade