Applications and Framework

Dr. Fayyaz Minhas

Department of Computer Science
University of Warwick
https://warwick.ac.uk/fac/sci/dcs/teaching/material/cs909/
Applications of Machine Learning

• An ability that I would like you to learn is to identify how to use machine learning in different domains.

• Machine learning can be applied in a wide array of real-world applications
Data, Big Data and Data Science

• We are going through an age of “Big Data”
  – Humans are the only biological entity that can store more data outside its body than inside it
Age of Big Data

Our Emergent Digital Future
Age of Big Data

Internet Transit Price (in Mbps)

Source: DrPeering.net

Internet users per 100 inhabitants
https://en.wikipedia.org/wiki/Global_Internet_usage
Moore’s Law – The number of transistors on integrated circuit chips (1971-2016)

Moore’s law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are strongly linked to Moore’s law.
Question

• Where does the bottleneck lie?
  – Data Analysis by Humans
  – Have humans become smarter?
  – How about we automate that?
Data Science

• Use of the scientific method in the development of:
  – Processes
  – Algorithms
  – Systems

• For extraction of knowledge from structured or unstructured data

• Examples
Linkedin’s Fastest-Growing Jobs Today Are In Data Science And Machine Learning

Louis Columbus, CONTRIBUTOR

Opinions expressed by Forbes Contributors are their own.

The world’s most valuable resource is no longer oil, but data

The data economy demands a new approach to antitrust rules

'Sexiest Job' Ignites Talent Wars as Demand for Data Scientists Soars
5th AVE NYC

1900

Where is the car?
5th AVE NYC

1913

Where is the horse?
Applications

US: 33 states accommodate self-driving vehicles on public roads.

Sweden: Last December, Volvo launched its Drive Me project, which provided self-driving cars to a number of people.

China: Shanghai issued its first self-driving licenses in 2018.

South Korea: The K-City is the largest town model ever built for self-driving car experimentation.

UK: The government passed a bill to draw up the liability and insurance policies related to autonomous vehicles.

California: In 2018, DMV allowed fully autonomous vehicles with no driver to operate on its public roads.

Arizona: Governor Ducey gave the green light for cars without drivers to operate on public roads in 2018.

Germany: the parliament passed a law last May that allows companies to test self-driving cars on public roads.

Netherlands: Council of Ministers first approved driverless vehicle road testing in 2015.

Singapore passed legislation recognizing motor vehicles don't require a human driver.

New Zealand: The country has no specific legal requirements for cars to have drivers.

Graphic: Tony Peng | Synced
From
Jim Elder
829 Loop Street, Apt 300
Allentown, New York 14707

To
Dr. Bob Grant
602 Quarenberry Parkway
Omar, West Virginia 2558

Nov 10, 1999

We were referred to you by Xena Cohen at the University Medical Center. This is regarding my friend, Kate Zack.

It all started around six months ago while attending the *Rotate* Jazz Concert. Organizing such an event is no picnic, and as President of the Alumni Association, a co-sponsor of the event, Kate was overworked. But she enjoyed her job, and did what was required of her with great zeal and enthusiasm.

However, the extra hours affected her health; halfway through the show she passed out. We rushed her to the hospital, and several questions, x-rays and blood tests later, were told it was just exhaustion.

Kate’s been in very bad health since. Could you kindly take a look at the results and give us your opinion?

Thank you!
Jim
The Letter
1978: First Postal Code Reader Worldwide
1982: First Address Reader Worldwide

1234 ELM STREET
Seattle, WA 98103
1984: First Multi Line Reader
1996: First Sender’s Address Reader

PAUL DRURY
1123 3RD AVENUE
NEW YORK, NY 10019

Mr. John Smith
1234 5th Street
Seattle, WA 98103
1998: First Full Text Reading
2000: First Graphics Recognition
2004: First Full Recognition
2008: Recognition on Both Sides of Envelope
OCR Accuracy

OCR Benchmarks
- NIST'93 Test Award
- ISRI'95 Award
- ICDAR'07 Arabic Award
Applications: Biometrics
Facebook Friends Tagging

We've Suggested Tags for Your Photos
We've automatically grouped together similar pictures and suggested the names of friends who might appear in them. This lets you quickly label your photos and notify friends who are in this album.

Tag Your Friends
This will quickly label your photos and notify the friends you tag. Learn more

Who is this?
Who is this?
Who is this?
Who is this?
Who is this?

Francis Luu
Skip Tapping Friends
Save Tags
Voice Assistants
Gmail: ML in NLP

William Shakespeare

Transformer AI poetry: Poetry classics as reimagined and rewritten by an artificial intelligence.

“Sonnet 18”

William Shakespeare

Shall I compare thee to a summer’s day?
Thou art more lovely and more temperate:
Rough winds do shake the darling buds of May,
And summer’s lease hath all too short a date;
A winter’s day, when beams of sun are few,
And skies are grey with clouds intemperate,
Is to thy beauty somehow sordid grown.
Her lovely-lovely eye is quite shut up,
Like windows pale before a sunny shower.
The earth that’s cold and moist, like to my mistress,
When once that she unbrack’d, her lovely jewel,
Grows as both ugly and uncourtly.
Poor soul, the center of my motley host,
See, now they request thy sovereign presence;
The din of merry hearts hath brought me thus
To greet thee, thou unsold and saint-like beauty.

More spam, but not in Gmail inboxes

- Spam prevalence: % of all incoming Gmail traffic (before filtering) that is spam
- Missed spam: % of total spam reported by Gmail users

As the amount of spam has increased, Gmail users have received less of it in their inboxes, reporting a rate less than 1%.
Recommender Systems

• Recommend movies based on user preferences, interests and likes
• Similar ideas for facebook...
  – Find friends that share your interests
Keyboard acoustics

Text recognized by the HMM classifier, with cepstrum features (underlined words are wrong),

the big money fight has drawn the shoporo od doses
of companies in the entertainment industry as well
as attorneys gnerals on states, who fear the
file sharing sofware will encourage illegal acyivitt,
grem the grosth of small arrists and lead to lost cobs and
dimished sales tax revenue.

Text after spelling correction using trigram decoding,

the big money fight has drawn the support of dozens
of companies in the entertainment industry as well as
attorneys generals in states, who fear the file sharing
software will encourage illegal activity, stem the growth
of small artists and lead to lost jobs and finished sales
tax revenue.

Original text. Notice that it actually contains two typographical errors, one of
which is fixed by our spelling corrector.

the big money fight has drawn the support of dozens
of companies in the entertainment industry as well as
attorneys gnerals in states, who fear the file sharing
software will encourage illegal activity, stem the growth
of small artists and lead to lost jobs and dimished sales
tax revenue.
PCR in HCI/CV

• Gesture Recognition
Synthesis
https://deepmind.com/blog/article/Agent57-Outperforming-the-human-Atari-benchmark
Google DeepMind's

AlphaFold 2

AI Breakthrough in Biology
Machine Learning

What my wife thinks I do

What I think I do

What I actually do

What society thinks I do

What my friends think I do

What other computer scientists think I do

\[
\min_\theta \frac{1}{2} ||\theta||^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i (\theta^T x_i))
\]

\[
h(\theta) = \frac{1}{2} ||\theta||^2 + C \sum_{i=1}^{n} \max(0, 1 - y_i (\theta^T x_i))
\]

Python 3.1.6 — An enhanced interactive Python.

- Introduction and overview of Python’s features,
- quickref — quick reference
- help — quick reference
- object — details about “object” use `object()” for extra details.

1. import Pytorch
2. import Pytorch

University of Warwick
Applied Machine Learning

• Examples from my research
Data Science: Hurricane Intensity Prediction

- Hurricane Intensity Prediction
  - In collaboration with National Hurricane Center, USA
  - Deep-PHURIE


Data Science: Journal Recommendation System

- Using classical NLP
- Using BERT
I can't be as confident about computer science as I can about biology. Biology easily has 500 years of exciting problems to work on. It's at that level.

— Donald Knuth —
COVID19 Meets Machine Learning
ECG Classification
Medical Data Classification
Predicting Protein interactions, interfaces and affinity

- Input: Two protein structures or sequences
- Output: What residue pairs interact

Experimental Validation: SnRK1- βC1

- In Silico Prediction and Validations of Domains Involved in Gossypium hirsutum SnRK1 Protein Interaction with Cotton leaf curl Multan betasatellite encoded βC1.

- βC1, pathogenicity determinant encoded by Cotton leaf curl Multan betasatellite interacts with calmodulin-like protein 11 (CML11) in Gossypium hirsutum.


In Silico Prediction and Validations of Domains Involved in Gossypium hirsutum SnRK1 Protein Interaction with Cotton leaf curl Multan betasatellite encoded βC1, Kamal, Hira, Fayyaz ul Amir Afsar Minhas, Hanu Pappu, Imran Amin et al., in Frontiers in Plant Science 10 (2019): 656.

Predicting anti-CRISPR proteins

- 20 proteins for training
- Identified 3 new anti-CRISPR proteins using a ranking ML model

Current Focus: PATHLake

- **PATHology data Lake, Analytics, Knowledge and Education**
- **UK Research and Innovation**
- **£15.7 million**
- **Objective:**
  - Improve speed and accuracy of cancer diagnosis
The Revolution in Pathology

Conventional Microscope Pathology

Digital Pathology

Computational Pathology
Why?

- Shortage of Pathologists
- Quantification is difficult
- Subjectivity
- Inter-observer variability
How much fat?
Steps in the development of a data science model

1. Define the Problem
   - Identify business goals
   - Identify data mining goals

2. Identify Required Data
   - Assess needed data
   - Collect and understand data

3. Prepare and Pre-process
   - Select required data
   - Cleanse/format data as necessary

4. Model the Data
   - Select algorithms
   - Build predictive models

5. Train and Test
   - Train the model with sample data sets
   - Test and iterate

6. Verify and Deploy
   - Verify final model
   - Prepare visualizations and deploy
Constructs of a Data Mining System for Prediction

• Identify the objective
  – Identify the unit of classification (example)
    • Image block, protein sequence, ....

Real world Phenomenon → Sensor → Feature extraction mechanism → Machine Learning → Decision
Learning from Data

• Example Case
  – Pathologists vs. Computer Scientists
    • Hypothetical!
    • Classify a person in their "native” environment
Constructs

• Sensor(s)
  – Camera

• Feature Extraction
  – White coats or lap aprons?
  – Computer Screen or microscopes?
  – Income?

• Machine Learning
Feature Space

Computer Screen?

Whiteness in Dressing

1/6 + 1/4 = 0.42

3/6 + 0/4 = 0.5

Whiteness in Dressing
Feature Space Classification

Computer Screen?

0/6 + 0/4 = 0.0

Whiteness in Dressing
So far?

- **Representation**
  - Represent examples in a feature space
  - Define a classification function
    - Line: \( f(x; w) = w_1 x^{(1)} + w_2 x^{(2)} + b = 0 \)

- **Evaluation**
  - Define an error function
    - Misclassifications

- **Optimize**
  - Reduce error

- **Real Test (Generalization)**
  - How does it perform on unseen data?

---

ML Output

Classifier

Feature Extraction

Input

Features

ML Output

Parameter/Weight Updates

Known Target of Training Example: +1
End of Lecture-2

We want to make a machine that will be proud of us.

- Danny Hillis