The Basics of Artificial Intelligence and Machine Learning

Alexis B. Carter, MD
Physician Informaticist and Molecular Genetic Pathologist
Presentation for CLIAC on April 10, 2024

Disclosures

In the past 12 months, I have not had any significant financial interest or other relationship with the manufacturers of the products or providers of the services that will be discussed in my presentation.

Having said that, attendees should be aware that:

- I am a paid faculty member of the AMIA Clinical Informatics Board Review Course.
- My spouse received consulting fees from Sysmex International, Inc.
- I am the immediate past Secretary/Treasurer for AMP
- I am on committees of several professional societies (CAP, AMP, CLSI) and one federal working group for CLIAC.

Goals and Objectives

- VERY high-level overview of Artificial Intelligence and Machine Learning (AI/ML)
- Describe current and future potential applications of AI/ML
  - Anatomic Pathology
  - Clinical Pathology
- Understand why it is critical for pathologists and laboratories to bring in data scientists to use AI/ML wisely

Artificial Intelligence (AI)

- Definitions
- Differences from traditional
- Uses / Benefits
- Challenges
- Published guidelines

Machine Learning (ML)

- Definitions (many)
- Learning & data terms
- Model & evaluation terms
- Quality metric methods
- Model development process
- Design, train, test, deploy

Machine Learning (ML) Algorithms

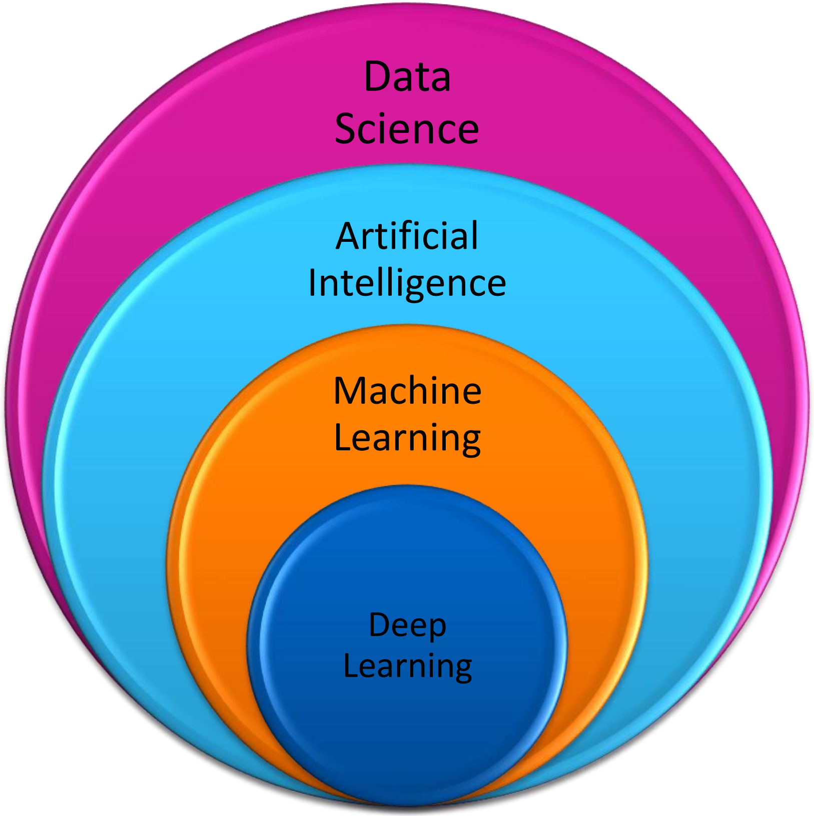
- Neural networks
- Supervised methods
- Regression, Classification, Ensemble
- Unsupervised methods
- Clustering, Association Rules, Dimensionality reduction

Artificial Intelligence

Definitions

Definitions

- **Data Science:** Science of organizing / analyzing massive amounts of data (In pathology = computational pathology)
- **Artificial intelligence (AI)**: ability of a computer or computer-controlled robot to perform tasks commonly associated with intelligent beings https://www.britannica.com/technology/artificial-intelligence
- **Machine Learning (ML):** Algorithms which allow computers to learn with**out** explicit programming
- **Deep Learning**: Specific set of ML tools designed to handle big data (e.g., specific neural networks)



Definitions

- **Narrow AI***
  - The machine can perform a **single** specific task better than a human
- **General AI**
  - The machine can perform **any intellectual task** with the **same** accuracy as a human
- **Strong AI**
  - The machine **outperforms** humans in **many** tasks

*** All currently deployed AI tools are only narrow AI.**

- "**AI Effect**" and "**Tesler's theorem**"
  - AI is whatever hasn't been done yet
  - Optical character and voice recognition, automated pap smear and peripheral blood smear readers, bioinformatics pipelines → no longer considered AI
- **Autonomous intelligence**
  - AI is making the decisions (no "human-in-the-loop")
- **Augmented intelligence**
  - AI is used to augment and/or assist humans in their work
  - Maintains "human-in-the-loop"; human ultimately making decisions

Why is Artificial Intelligence different?

Machine Learning vs. Traditional Programming

Traditional Programming

Input Data

Human-specified **rules** analyze data according to **known** or **suspected** patterns

Output Data

Machine Learning

Input and **output** data

Machine Learning Model / Tool

Human does not have to write code or even know the patterns for analysis to develop the algorithm

Computer-specified **rules** for analysis

Machine Learning vs. Traditional Statistics

Function	**Traditional Statistics**	**Machine Learning**
Defines explicit mathematical relationship between inputs and outputs	Yes	Not usually
Makes assumptions about characteristics and distribution of the data fed to it •Parametric vs. Non-parametric •Normal distribution vs. Non-normal distribution	Yes	Not usually
Handles large # input variables	Not usually	Yes
Can use complex multifactorial data	Not usually	Yes
Reason for output is clear and explainable	Yes	Not usually (**black box problem**)

Uses and Benefits of Artificial Intelligence and Machine Learning (AI/ML)

Uses and Benefits – Anatomic Pathology

- Classifications
  - Current Hype
    - Histopathologic diagnosis through image analysis (active research area)
  - Actual current and possible uses
    - Smart assistive technology for pathologists to make diagnoses better, faster
      - Counting mitoses
      - Finding tiny metastases
      - Detecting sneaky microorganisms
- Predictions based on histologic features
  - Prognosis of patient
  - Molecular sub-characterization
- Anomaly detection
  - Detecting errors in data (e.g., pathology reports…Ye JJ, Tan MR, *J Pathol Inform*, 2019; 10:20)

Uses and Benefits – Clinical Pathology

- Predictions
  - Lurking medical diagnoses from general laboratory test results (e.g., future anemia from CBC trends)
  - Patient volumes → adjust staffing
  - Determination of optimal future state workflows / functional gaps in process redesign
  - Predicting, detecting and subverting malware attacks
- Classifications
  - Pattern detection (e.g., diagnoses), feature detection (images)
  - NGS variant pathogenicity algorithms
  - Variant prioritization of variants determined through exomes and genomes
- Decision support
  - Making prior authorization decisions
- Signal conversion
  - E.g., natural language processing, voice recognition, optical character recognition
- Anomaly detection
  - Problem-solving for unexpected laboratory results
  - Monitoring for shifts and trends in live result data that may indicate instrument problem before the next QC run

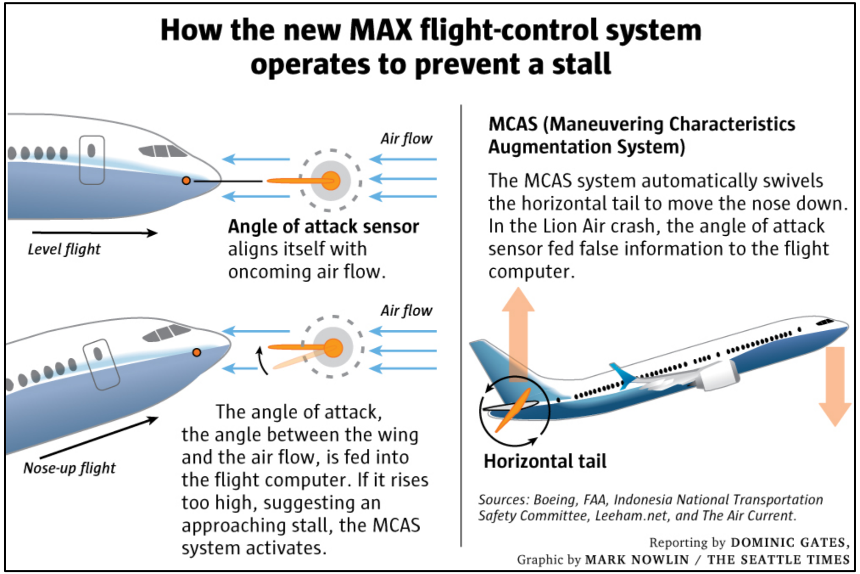
Challenges

Challenges

- Some challenges similar to other non-AI software
  - Cybersecurity risks
  - Software can be developed with bad data or bad science
  - **Automation bias** – assumption that the computer is right, even when it doesn't make sense [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3240751/]
  - Inaccurate assumptions about data accuracy and representation
    - https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care

Challenges Illustrated - Story of Harm

- **Boeing 737 MAX flight control system**
  - Two plane crashes killing all 346 passengers in Oct 2018, Mar 2019
  - **Faulty** angle-of-attack **sensors** fed bad data to system
  - **No redundant sensors** required to detect when sensor was faulty
  - **No usable human override mechanism**
  - **Default configuration did not show alerts** for mismatched sensor data (when >1 sensor present)
  - **System was not set to disengage** when multiple errors generated at once
  - **Similar errors during simulations not reported to FAA** by Boeing because they were considered "advisory" rather than "critical"
  - FAA, citing lack of funding and resources, over the years had delegated increasing authority to Boeing to assess its own work during certification processes



- Image from: https://arffwg.org/max-737-sensor-w/
- https://arffwg.org/max-737-sensor-w/
- https://www.washingtonpost.com/transportation/2019/05/15/faa-chief-be-pressed-boeing-max-while-would-be-replacement-faces-questions-his-approach-air-safety/?noredirect=on&utm_term=.ffb046749452
- https://www.faa.gov/foia/electronic_reading_room/boeing_reading_room/media/737_RTS_Summary.pdf

Challenges – Data Quality

- Good quality data is **critical**
  - bad data → bad model
  - Some models need large amount of training data
- Data have insufficient quantity / variability for context
  - Especially problematic for models finding less common patterns (e.g., disease screening, anomaly detection)
  - Underrepresented populations → non-generalizable rules (socioeconomic, gender, race, ethnic and other disparities)
- Data labels represent human bias / false beliefs
  - e.g., court sentences, hiring / firing decisions
  - Can promulgate or exacerbate inequality
- Data have incomplete, inaccurate and/or variable labels
  - Different terms or metrics for same label due to human inconsistency
- Critical input data may be missing
  - **Polanyi's Paradox**:
    - Human decision-making beyond explicit understanding or description
  - Human may not realize which data contributed to human decision
  - Critical inputs may not be represented in AI training data

Challenges – ML Model Problems

- Models can be brittle
  - Small changes in input → big changes in output
  - Unable to see the forest for the trees (double-edged sword)
  - Humans are BETTER at generalization and situational awareness
- Small changes to input introduced by hackers (**adversarial examples**) led to wrong output [https://www.nature.com/articles/d41586-019-03013-5]
- Models can also degrade over time
  - Similar concept for laboratory tests (drift, shift)

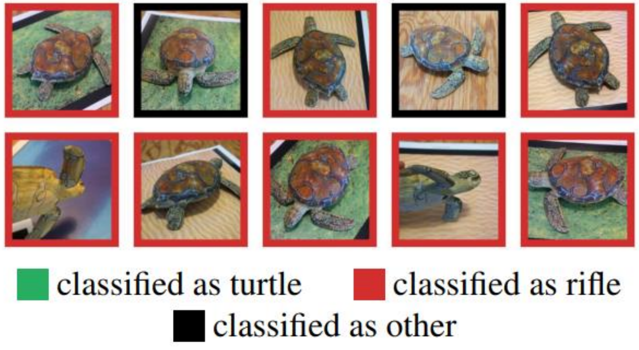


Figure 1. Randomly sampled poses of a 3D-printed turtle adversarially perturbed to classify as a rifle at every viewpoint[2]. An unperturbed model is classified correctly as a turtle nearly 100% of the time.

Athalye et al. 2018.
https://arxiv.org/pdf/1707.07397.pdf

Challenges - Cybersecurity

- AI can be hacked just like any other software
  - Robotic surgical systems (https://www.ncbi.nlm.nih.gov/pubmed/30397993)
- Hacked systems have potential for unauthorized disclosure, patient harm
- Human autonomy ("human-in-the-loop") may help detect malfunctions
- US national efforts for AI cybersecurity
  - National Security Commission on Artificial Intelligence (https://www.nscai.gov/)
    - Established 2018 by John S. McCain National Defense Authorization Act (Public Law 115-232)

Challenges - Transparency

- Definitions (multiple)
  - For AI developers: Reasons for model's performance are **known** and **understood**
  - For end-users (ethics): Sufficient information is published such that model's performance can be audited [https://www.who.int/publications/i/item/9789240029200]
- Lack of transparency (**Black box problem**)
  - Rules developed by the AI algorithm
    - May be indecipherable after model is trained, even to the developer(s)
    - May not be able to determine why algorithm generated certain output
    - May generally work well but some output may be inexplicably wrong

Challenges - Ethics

- Hot topic because of some noted failures
  - https://georgetownsecuritystudiesreview.org/2021/05/06/racism-is-systemic-in-artificial-intelligence-systems-too/
  - https://technologyandsociety.org/bias-and-discrimination-in-ai-a-cross-disciplinary-perspective/
  - https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/
- **Beneficence:** Maximize benefits; minimize risks and harms
  - AI can propagate and exacerbate human bias
  - Protect human autonomy in decisions (**"human-in-the-loop"**)
    - ACR and RSNA recommendation → do not approve autonomous AI until sufficient human-supervised AI experience obtained
- **Auditability**: Audit the tool to verify performance, ensure ethics followed
- **Accountability**: Who or what is accountable when something goes wrong
  - Medicolegal liability
    - AI is not standard of care
    - Regulations not yet developed in US
    - EU paper (https://pubmed.ncbi.nlm.nih.gov/33489979/) that discusses that liability is based on physician using standard of care

Challenges – Ethics (cont.)

- **Intelligibility**
  - Achieved through Transparency and eXplainability
    - https://nvlpubs.nist.gov/nistpubs/ir/2020/NIST.IR.8312-draft.pdf
  - **Transparency** [https://www.who.int/publications/i/item/9789240029200]
    - Sufficient information **published** before the design or deployment of an AI technology
      - Describes how technology is designed, intended use, data used, etc.
    - Also means that a person knows when AI is being used on them
  - **eXplainability** (XAI)
    - Providing the human user an explanation of how the AI tool works

Other Challenges

Personnel

- Medicine lacks sufficient data scientists
- Many data scientists lack expertise in medicine and/or healthcare environment

Organizational

- Lack AI strategies
  - Right tasks
  - Right data
  - Right evidence standard(s)
  - Right approaches for integration
- Deploying models in clinical environments is challenging (patient safety, population differences between locations)

Financial

- Lack of reimbursement mechanisms
- Harder to define returns on investment

Technical

- Lack of adequate computational infrastructure
- Introduces new cybersecurity threats that aren't yet addressed

Response to Challenges → Guidelines

- Guideline for machine learning model development (US, Canada, UK Guideline – Oct 2021)
  - https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles
  - Multidisciplinary expertise throughout
  - Good software/security practices
  - Data representative of intended patient population
  - Training data independent of testing data
  - Reference data is well characterized
  - Model design tailored to available data and reflects intended use
  - Focus on keeping the human in the loop (human AI team)
  - Testing demonstrates performance during clinically relevant conditions
  - Users provided clear essential information for use
  - Deployed models are monitored for performance in the real world
- AI Ethics Guidelines and White Papers
  - WHO Ethics Guidelines for AI https://www.who.int/publications/i/item/9789240029200
  - UNESCO https://unesdoc.unesco.org/ark:/48223/pf0000379920.page=14
  - EU guidelines https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
  - https://www.intelligence.gov/artificial-intelligence-ethics-framework-for-the-intelligence-community

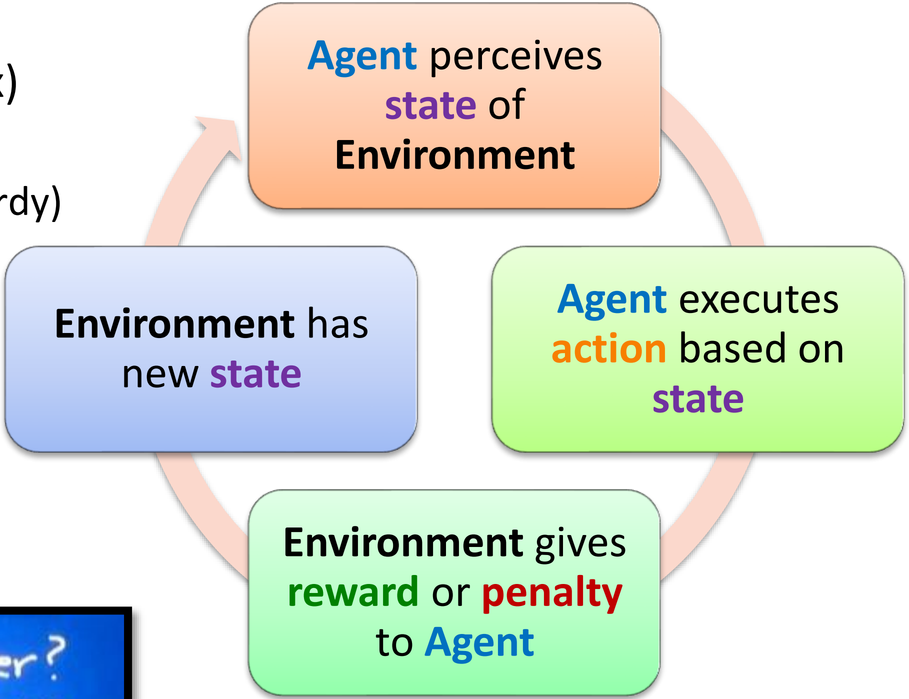
Machine Learning Rudimentary Basics

ML Definitions – Types of Learning

Supervised learning	Trains on classified and/or labeled data • Goal → train model to generate **known** answers, patterns or relationships
Fully supervised	All data labeled to same extent (degree of detail)
Semi-supervised	Some data are labeled while other data are not • Unlabeled data may be auto-labeled to match patterns on labeled data
Weakly supervised	Small amount of data have detailed labels; rest of data have fewer labels
Unsupervised learning	Data which have **not** been classified or labeled • Goal → model discovers **new** (previously **unknown**) patterns or relationships

ML Definitions – Types of Learning

- **Reinforcement learning**
  - Used to learn how to reach a (complex) goal
    - Game playing (IBM Watson and Jeopardy)
    - Speech to text, financial trading



ML Definitions – Types of Learning

- **Transfer learning**
  - Separate category vs. subtype of supervised learning
  - Data used for training the model are transferred from a different related domain
    - Data were developed for use in a domain different than the one intended for the model
    - Example: Using natural images from ImageNet (https://image-net.org/) to train a models for medical images [Alzubaidi et al 2021 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8036379/]
  - Coarse training done on transferred data
  - Fine tune training with smaller data directly related to domain of use
  - Reasons
    - Data are expensive
    - Higher quality and quantity data may be more available, cheaper in another domain

ML Definitions - Data

- **Instance**
  - Single event in a data set
  - # instances required to train a model depends on the problem and model used
  - **Outlier**
    - Instance which is significantly different from the remaining instances in the population
    - Can skew results
    - Different models have different sensitivities to outliers
- **Label** – observed value for a feature of an individual instance
- **Feature**
  - An aspect (variable) of the training data
  - Called a **dimension** in unsupervised learning

	Feature 1	Feature 2	Feature 3
Instance 1	Red	Slow	Yes
Instance 2	Red	Fast	No
Instance 3	Green	Medium	No

Red, Green, Slow, Fast, Medium, Yes and No are all **labels** in this data set.

ML Definitions - Models

- **Algorithm**
  - Repeatable process used to train a model from a given set of training data
- **Parameter**
  - Internal values inside machine learning that the model derives based on training data
  - e.g., weights, bias values
- **Model** = algorithm + parameters
  - When a model is used for classification, it is called a **classifier** [https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623]
  - **Weak learner (weak model):** model whose performance only slightly > random chance
  - Good model: model that **generalizes well** (it performs the same on new data as it did on the training (and test) data)
- **Epoch**
  - 1 epoch = 1 pass through the training data

ML Definitions – Model Evaluation

Signal

The true underlying pattern you are trying to learn from the data

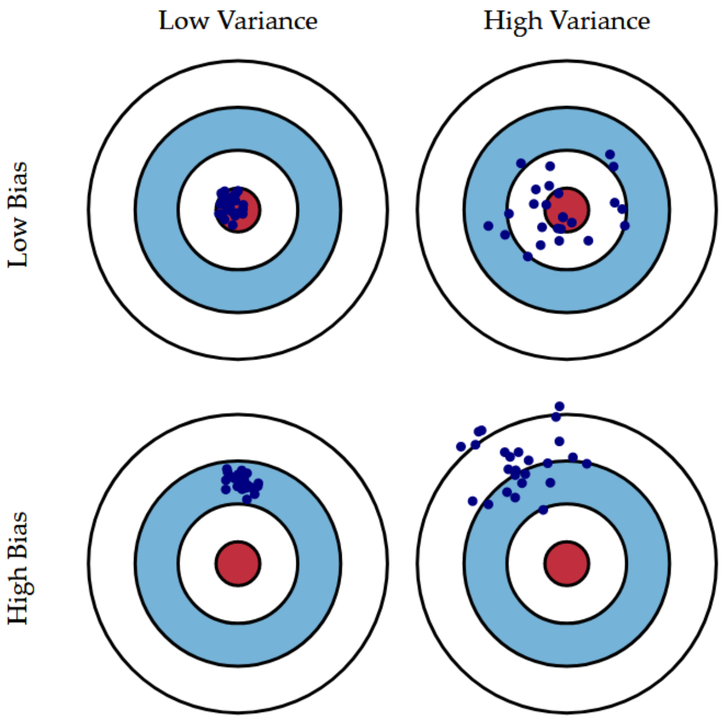
Well designed machine learning separates signal from noise

Noise

Irrelevant information or randomness in a data set

Irreducible error

Bias	Variance	Irreducible error
• Measure of inaccuracy • High bias + low variance ➔ consistently inaccurate results	• Measure of imprecision (lack of reproducibility) • High variance + low bias ➔ inconsistently accurate results	• Noise that cannot be reduced by optimizing algorithms



https://devopedia.org/bias-variance-trade-off

ML Definitions – Model Evaluation

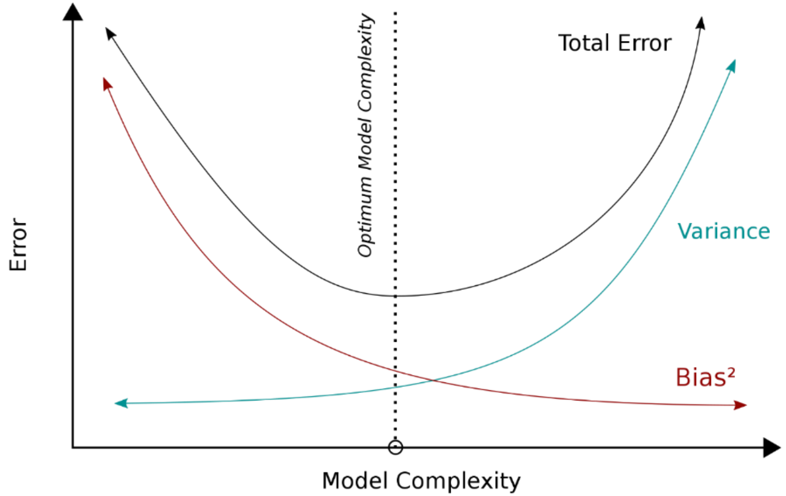
Bias

- *Not just an ethical term…*
- Amount of **inaccuracy** in the model's performance after training
- High bias → model is inaccurate (underfit)
- Low bias → model is accurate (but may be overfit)

Variance

- Amount of **imprecision** (square of standard deviation (σ) → σ^2)
- Due to model's sensitivity to small fluctuations in the training set
- High variance → model is imprecise (and likely overfit)
- Low variance → model is precise (but may not be accurate and may be underfit)

ML Definitions – Model Evaluation



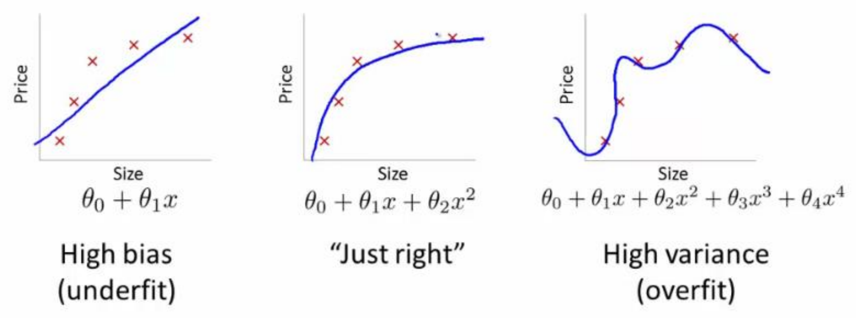
- **Bias-Variance Trade-Off**
  - Things that reduce variance increase bias
  - Things that reduce bias increase variance

$$Total\ error = (bias^2) + variance + irreducible\ error$$

https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff
https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

ML Definitions – Model Evaluation

- **Goodness of fit**
  - How closely a model's output values match the observed (true) values
- **Underfitting**
  - Model does not accurately predict output for the data fed to it
    - high bias, low or high variance
- **Overfitting**
  - Occurs when statistical model exactly fits training data BUT…
    - Does not fit new data well (test or production data)
  - Training set has low error rate but test set has high error rate = high variance
  - **Most common problem** for any statistical model using a training set



https://datascience.stackexchange.com/questions/361/when-is-a-model-underfitted

ML Definitions – Model Evaluation

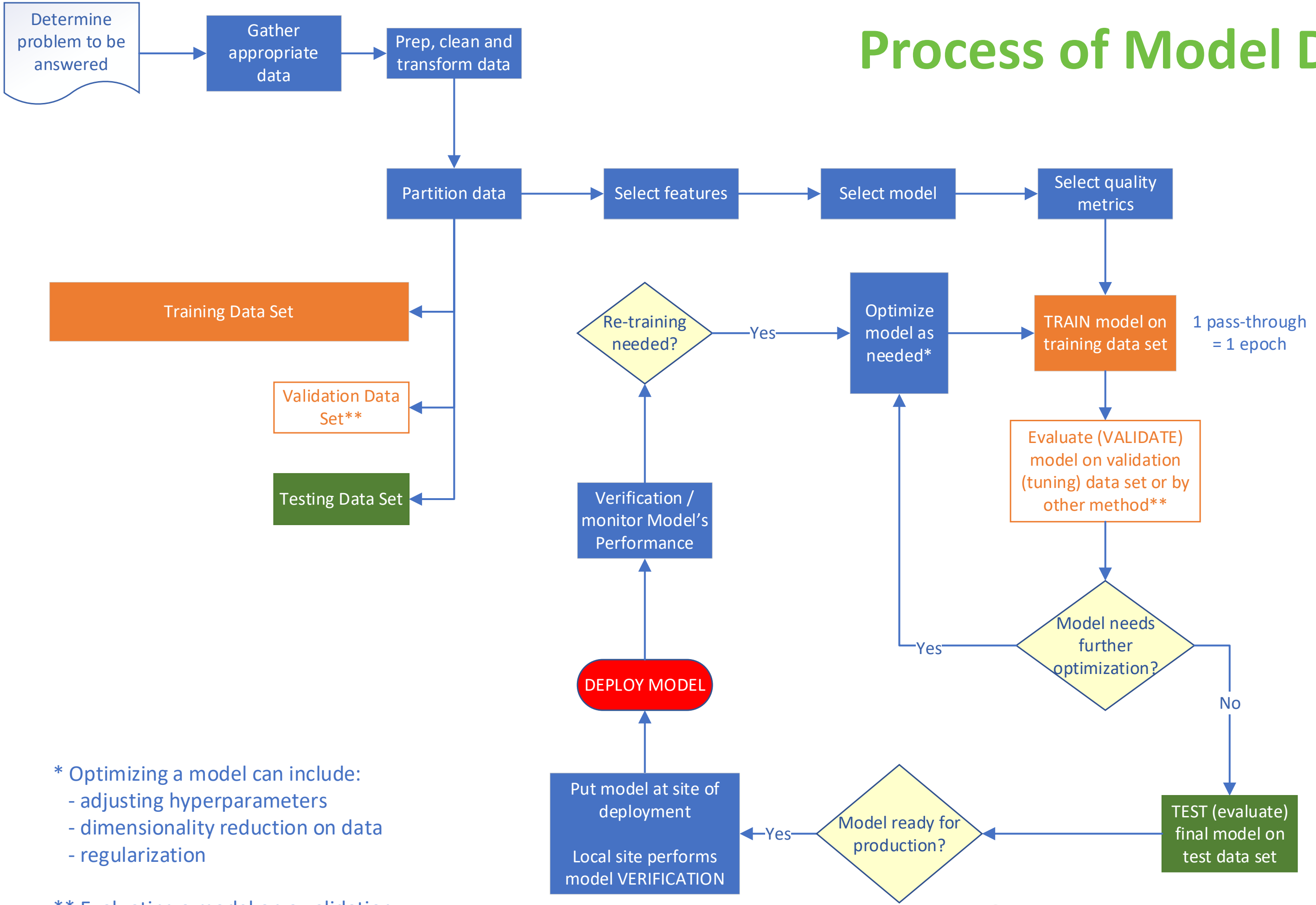
- **Null error rate**
  - For classification methods, rate of being <u>wrong</u> if you ALWAYS pick the majority class
  - If the majority class has 105 instances out of 165 total instances
    - Null error rate = (165 – 105)/165 = 36%
  - **Accuracy paradox**
    - Best classifier for the intended use may have a higher error rate than the null error rate
    - Occurs when condition or outcome is very low percentage of overall data set (e.g., 1%)
    - Model can correctly predict absence of the condition in 99% of cases – hooray! BUT…
    - May completely fail to detect the condition being sought
      - 100% failure of detecting the condition (but null error rate is only 1%)
    - Take home point → Use different statistical methods when trying to screen for low incidence conditions

Process of ML Model Development

- Many ways that a model can be trained → tested → deployed
  - Depends on model, amount of data, and other factors
- Phases of model development have variable nomenclature between authors
  - E.g., learning phase, inference phase
- A few definitions to resolve possible confusion

	What it means in machine learning…	**What it means in a hospital laboratory…**
Validation	Evaluating preliminary (non-final) *model* • Results of evaluation lead to tweaking (tuning) the model	Final evaluation of a *laboratory test* where no further changes to the test procedure are expected
Testing	Final evaluation of a *machine learning model* where no further changes to the model are expected	Evaluating preliminary (non-final) *laboratory test* OR Performing live clinical testing

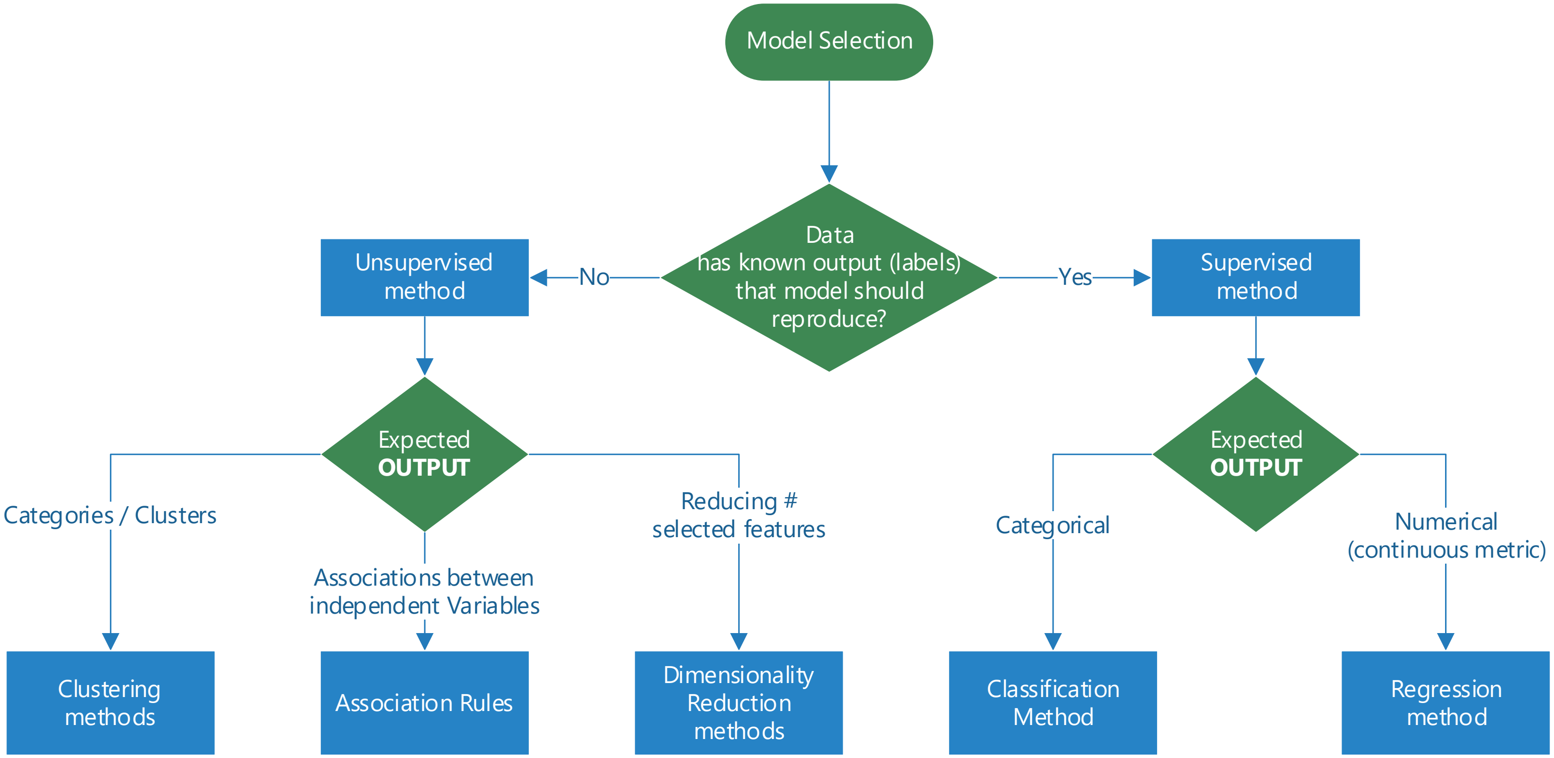
Process of Model Development



* Optimizing a model can include:
- adjusting hyperparameters
- dimensionality reduction on data
- regularization

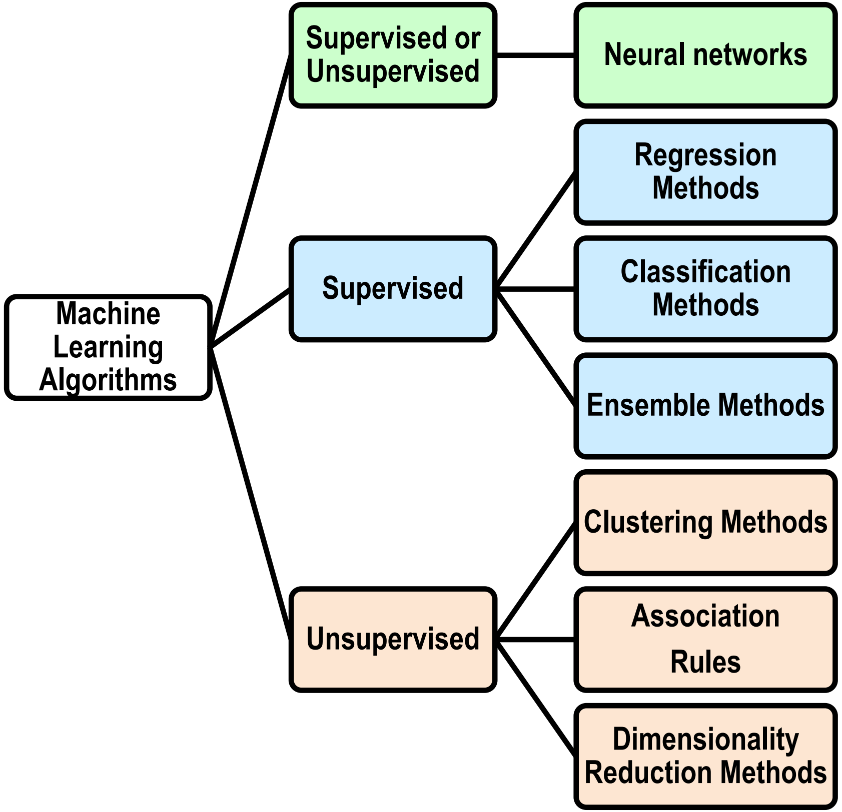
** Evaluating a model on a validation data set may not always be needed.

Process of Model Development



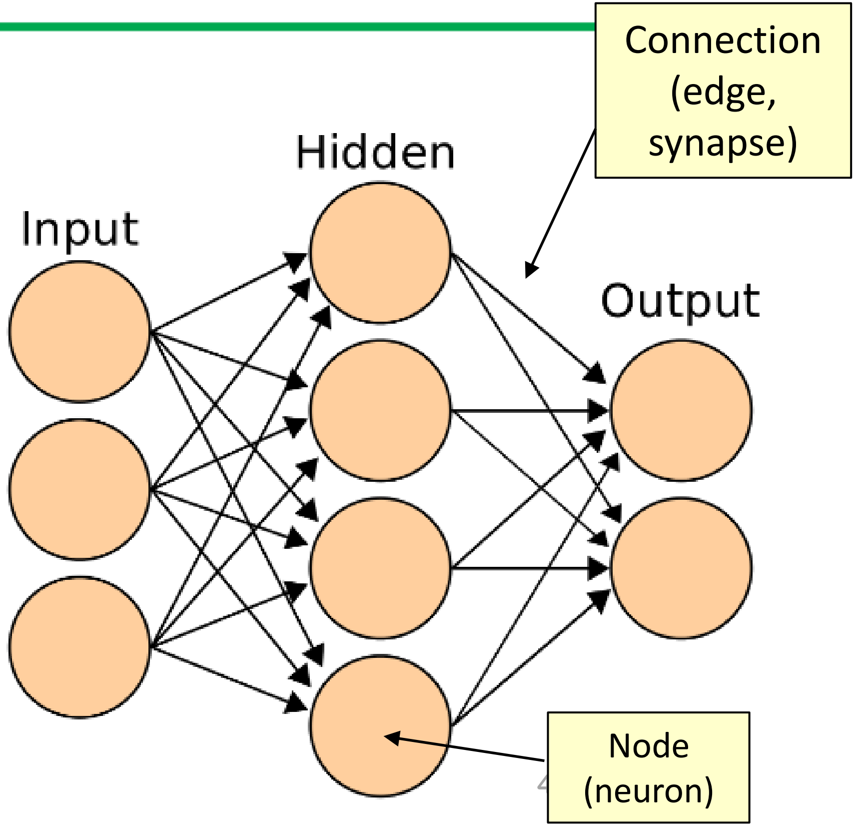
Machine Learning Algorithms

- Each category has algorithms that are primarily used for that purpose
- However, classification algorithms may sometimes be used for regression and vice versa
- Unsupervised algorithms may sometimes be used with supervised learning



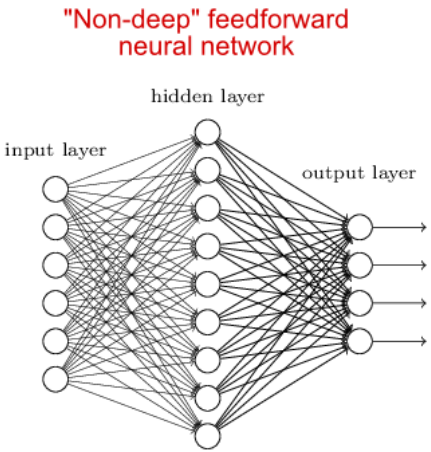
Artificial Neural Networks (ANNs)

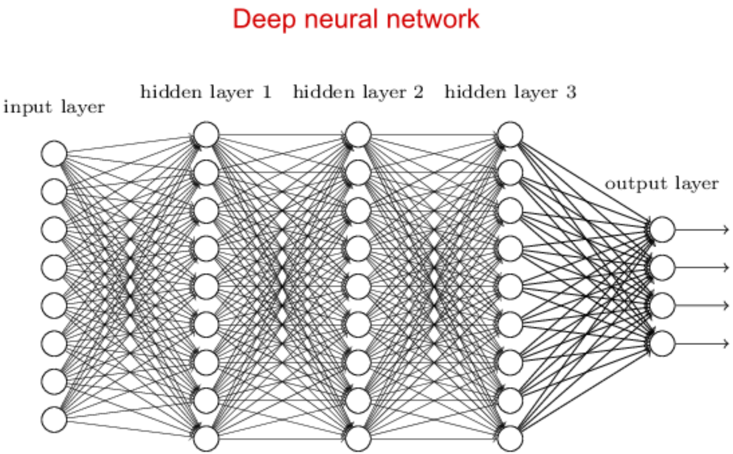
- Goal: Solve problems like a human
- Operate via flow through neural nets, akin to biological networks
  - Handles large amounts of complex data
  - Computationally intensive
  - Unraveling the pathways after training is completed can be difficult to impossible → **Black Box Problem**
- **Nodes** (akin to neurons) → transfer functions
- **Connections** (akin to synapses, a.k.a. edges)
- **Back-propagation** (nice YouTube (https://www.youtube.com/watch?v=Ilg3gGewQ5U) video)
  - Learns mistakes based on output
- Layers (nodes in each layer *usually* have same activation function)
  - **Input layer**: # nodes = # features selected in data
  - **Output layer**: # nodes = # output categories of data
  - **Hidden layer(s): Shallow networks** usually have 1; **Deep networks** have >3



ANN – Deep Learning

- **Deep Learning** (a.k.a. deep networks; deep nets)
  - Goal: *imitate the human brain* in processing data and decision-making patterns
  - Usually multiple (Some say > 1 to >3 to hundreds to thousands) of hidden layers
    - Thousands to millions of interconnections; large number non-linear computations
  - Means more in-depth processing, *not* more in-depth knowledge





https://stats.stackexchange.com/questions/182734/what-is-the-difference-between-a-neural-network-and-a-deep-neural-network-and-w

References

- Hoyt R, Muenchen R. *Introduction to Biomedical Data Science*. Lulu.com, 2019. https://www.informaticseducation.org/about-the-textbook.
- Bowers D. *Medical Statistics from Scratch: an Introduction for Health Professionals.* Hoboken NJ: Wiley Blackwell, 2014.
- Callaway J. *Machine Learning: the Ultimate Guide.* Columbia, SC: 2021. ISBN 9798750222902.
- Burkov A. *The Hundred-Page Machine Learning Book.* 2019. ISBN 978-1-9995795-0-0.
- https://www.nist.gov/project-category/materials-genome-initiative-mgi/machine-learning-ai
- https://www.nist.gov/artificial-intelligence
- https://towardsdatascience.com/which-machine-learning-model-to-use-db5fdf37f3dd
- https://blogs.sas.com/content/subconsciousmusings/2020/12/09/machine-learning-algorithm-use/

Questions?