The Basics of Artificial Intelligence and Machine Learning

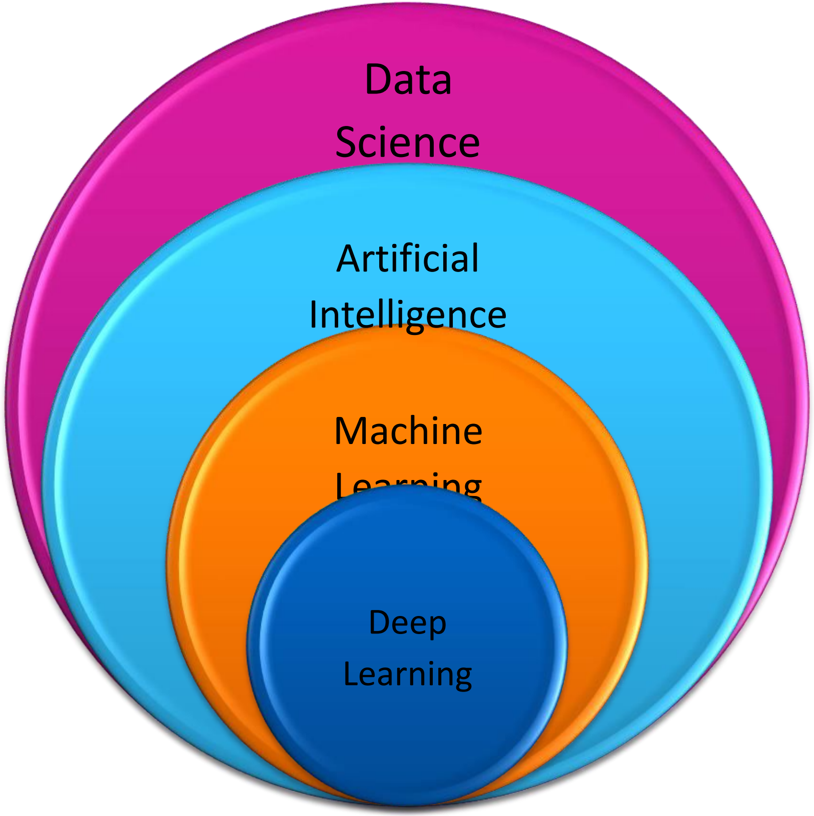
~~GM

Artificial Intelligence

Definitions

Definitions

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



Definitions

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

1. Siri, Alexa, or Google Assistant for specific tasks.
2. Facial recognition systems in smartphones.
3. Netflix or Amazon recommendation engines.

1. AI that learns and performs tasks across diverse fields.
2. A robot capable of diagnosing diseases and writing novels.

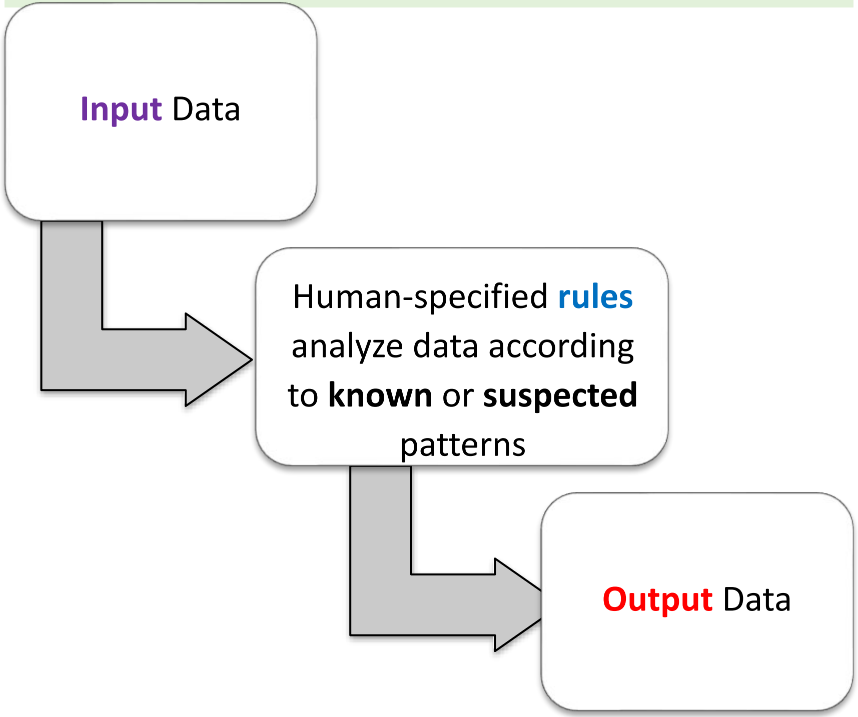
1. Robot surgeons outperforming human doctors.
2. AI managing corporations or governments with superior efficiency.

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

Why is Artificial Intelligence different?

Machine Learning vs. Traditional Programming

Traditional Programming

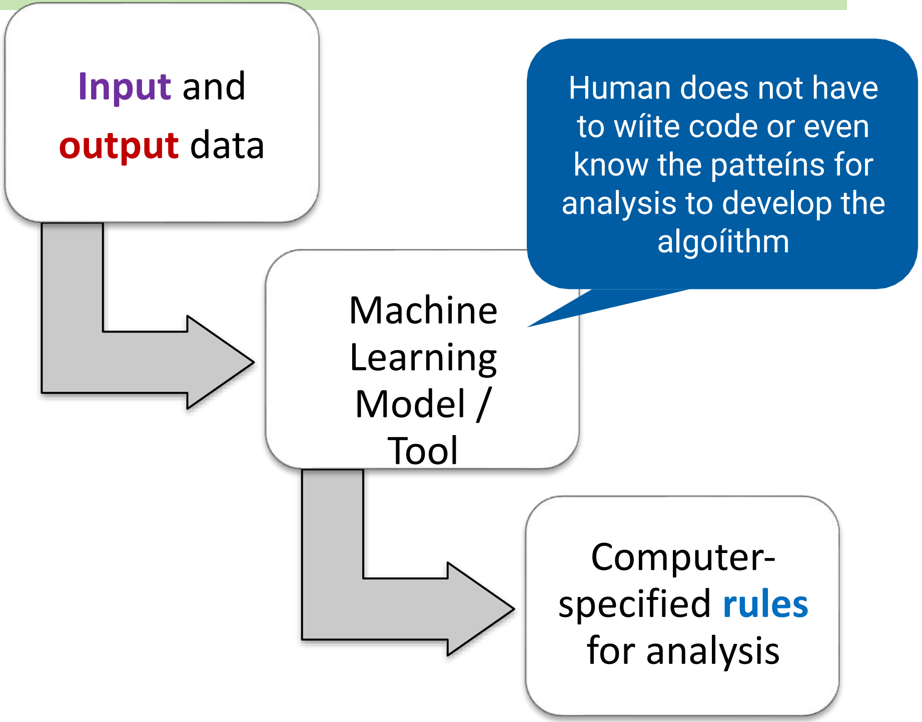


Input Data

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 wíite code or even know the patteíns for analysis to develop the algoíithm

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)

AI
A
AI
AI
AI
AI
AI

Healthcare

1. Drug discovery and personalized medicine.
2. Disease diagnosis (e.g., cancer detection).

Medical Imaging and Diagnostics

1. AI-powered tools analyze X-rays, MRIs, and CT scans for faster and more accurate diagnosis.
2. Early detection of diseases like cancer, tumors, and heart conditions.

Finance:

1. Fraud detection and risk assessment.
2. Algorithmic trading and customer insights.

Transportation:

1. Autonomous vehicles (self-driving cars).
2. Route optimization and traffic prediction.

Retail:

1. Personalized recommendation.
2. Inventory management and demand forecasting.

Manufacturing:

1. Predictive maintenance for machinery.
2. Quality control using visual inspections.

Education:

1. Adaptive learning platforms.
2. Automated grading and personalized tutoring.

Entertainment:

1. Personalized recommendation.
2. Inventory management and demand forecasting.

Examples

1.CleverBot

2.Autonomous Cars

3.Drones

4.Watson

AI Controversies

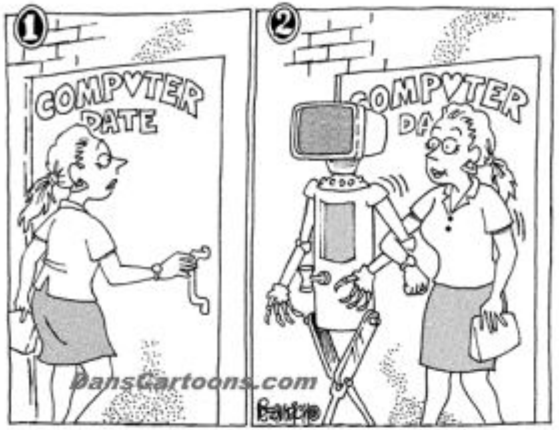
Potential job takeover

Growing laziness

Growing Cost

Priority Argument

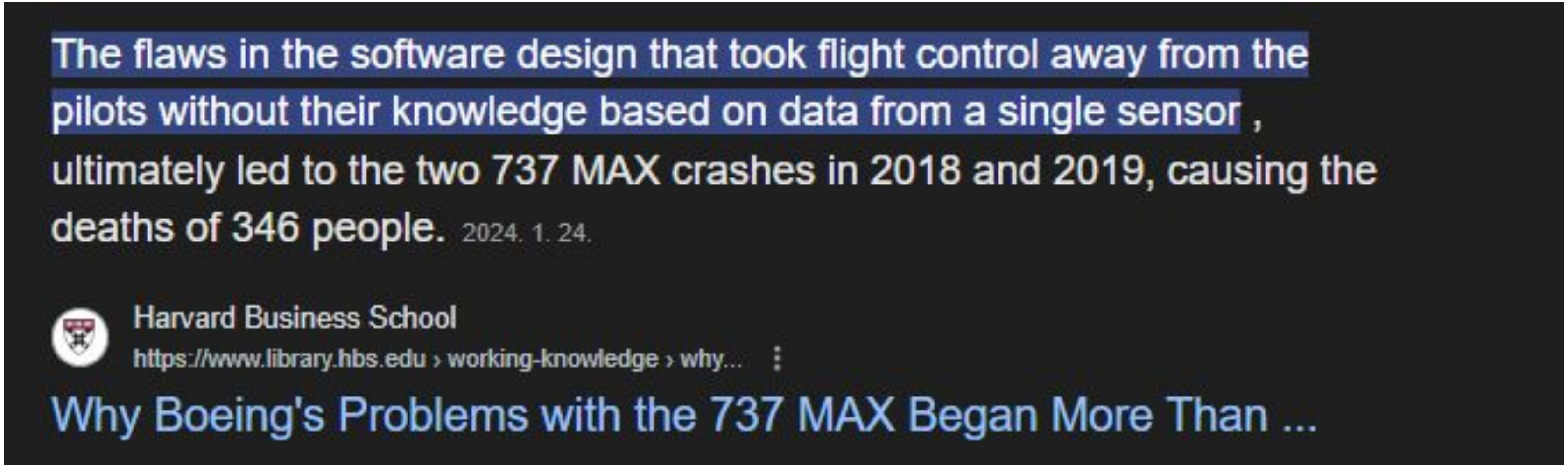
Robot Relationships?



Challenges

Challenges

- Some challenges similar to other non-AI software
  1. Cybersecurity risks.
  2. Software can be developed with bad data or bad science.
  3. **Automation bias** – assumption that the computer is right, even when it doesn't make sense.
  4. Inaccurate assumptions about data accuracy and representation.



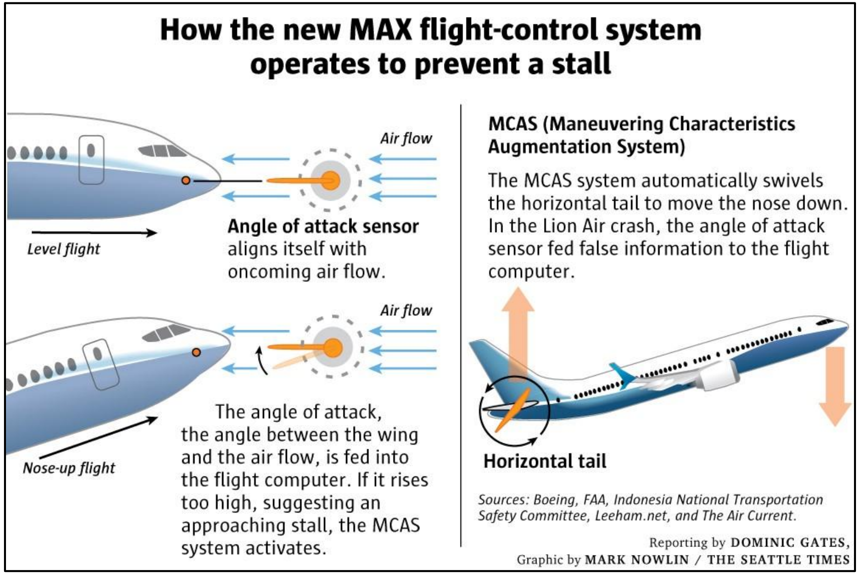
The flaws in the software design that took flight control away from the pilots without their knowledge based on data from a single sensor , ultimately led to the two 737 MAX crashes in 2018 and 2019, causing the deaths of 346 people. 2024. 1. 24.

Harvard Business School
https://www.library.hbs.edu › working-knowledge › why...

Why Boeing's Problems with the 737 MAX Began More Than ...

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



How the new MAX flight-control system operates to prevent a stall

Angle of attack sensor aligns itself with oncoming air flow.

The angle of attack, the angle between the wing and the air flow, is fed into the flight computer. If it rises too high, suggesting an approaching stall, the MCAS system activates.

MCAS (Maneuvering Characteristics Augmentation System)

The MCAS system automatically swivels the horizontal tail to move the nose down. In the Lion Air crash, the angle of attack sensor fed false information to the flight computer.

Sources: Boeing, FAA, Indonesia National Transportation Safety Committee, Leeham.net, and The Air Current.

Reporting by DOMINIC GATES,
Graphic by MARK NOWLIN / THE SEATTLE TIMES

- 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)

- Data have incomplete, inaccurate and/or variable labels
  - Different terms or metrics for same label due to human inconsistency

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]
  - Similar concept for laboratory tests
- Models can also degrade over time (drift, shift)

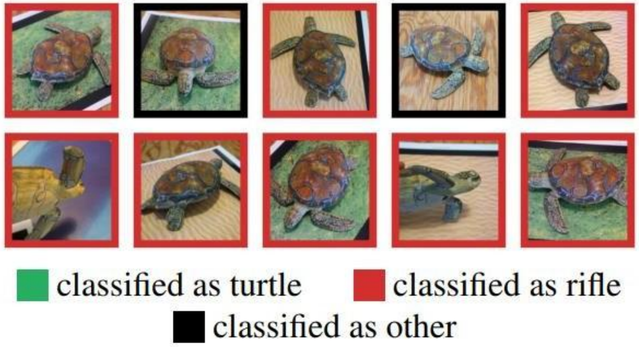


Figure 1. Randomly sampled poses of a 3D-printed turtle adversarially perturbed to classify as a rifle at every viewpoint[2]. An unperturbed model is classified correctly as a turtle nearly 100% of the time.

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)

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

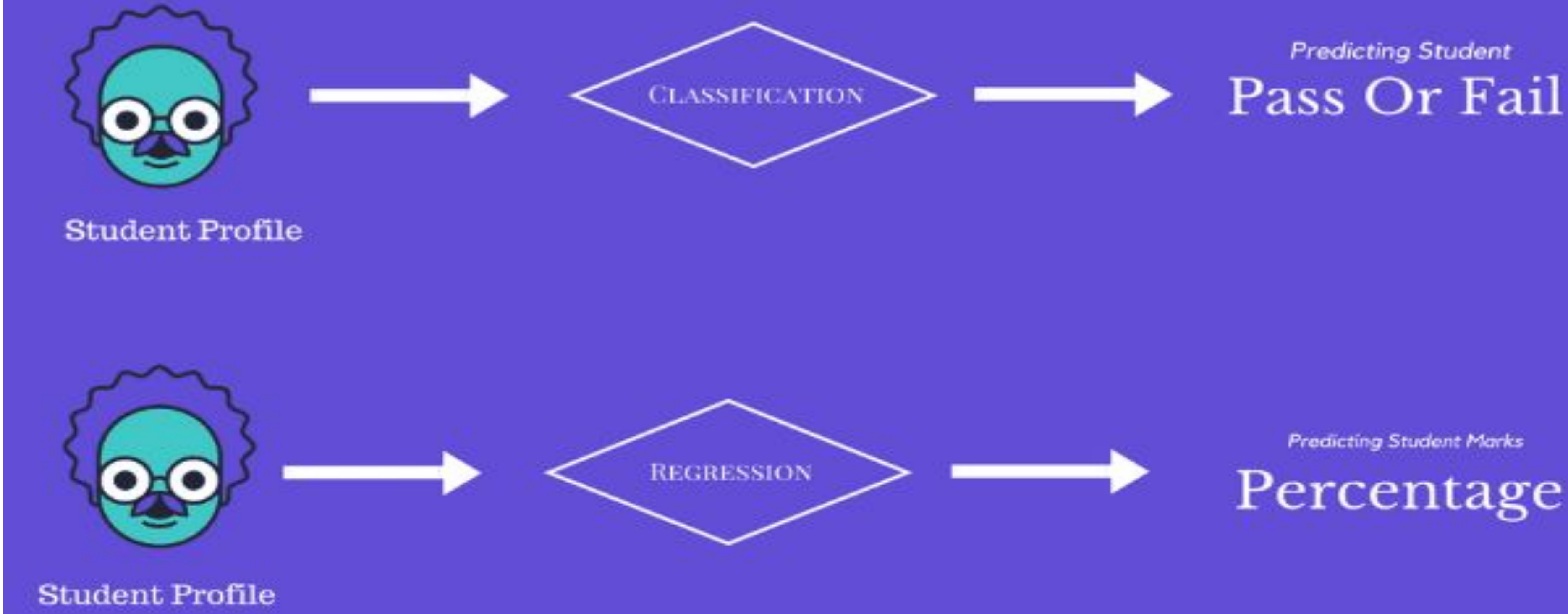
Technical

- 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

CLASSIFICATION VS REGRESSION



Student Profile

CLASSIFICATION

Predicting Student

Pass Or Fail

Student Profile

REGRESSION

Predicting Student Marks

Percentage

Classification

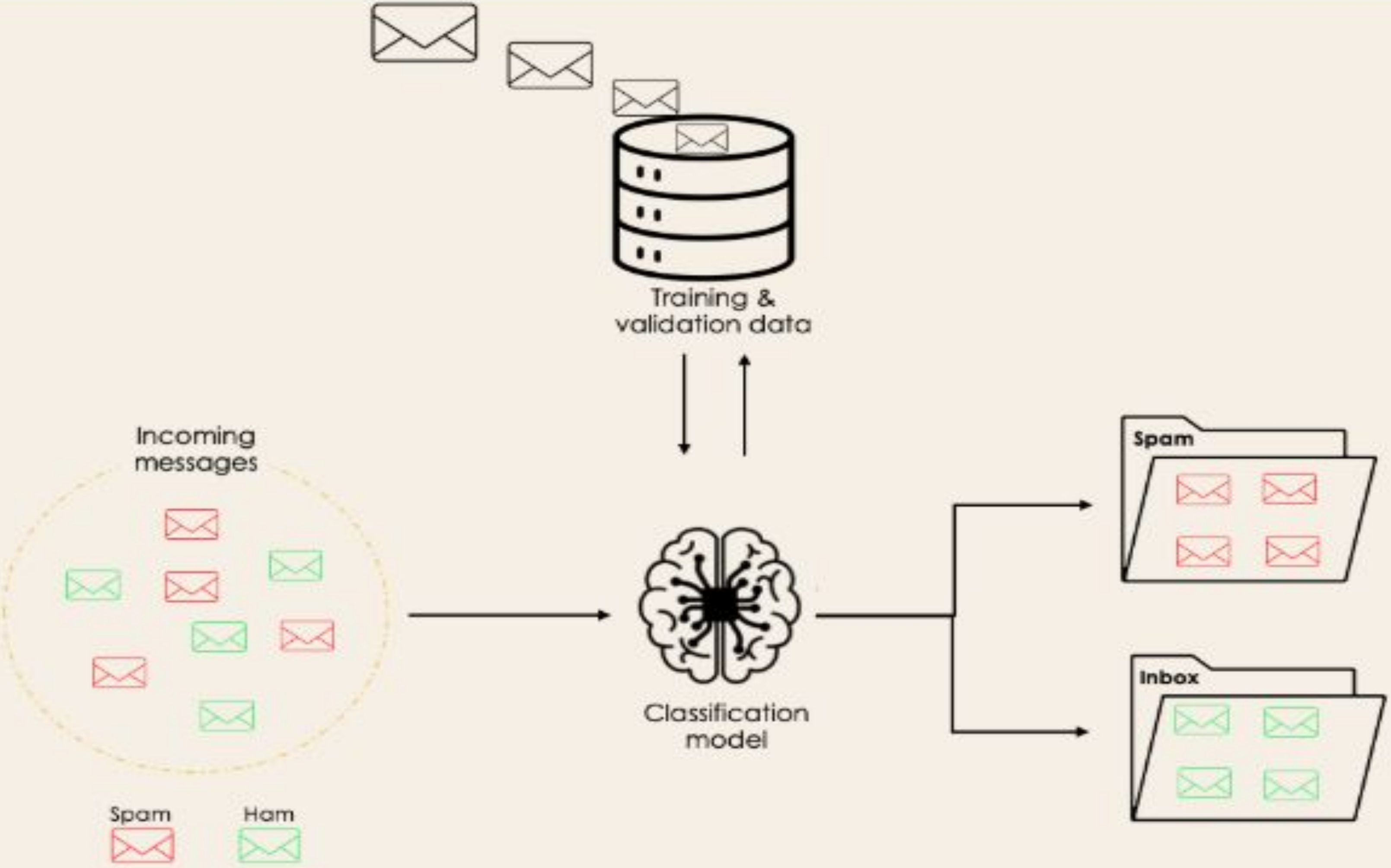
The main goal of classification is to predict the target class (Yes/ No). If the trained model is for predicting any of two target classes. It is known as binary classification. Considering the student profile to predict whether the student will pass or fail. Considering the customer, transaction details to predict whether he will buy the new product or not. These kind problems will be addressed with binary classification. If we have to predict more the two target classes it is known as multi-classification. Considering all subject details of a student to predict which subject the student will score more. Identifying the object in an image. These kind problems are known as multi-classification problems.

Regression

The main goal of regression algorithms is the predict the discrete or a continues value. In some cases, the predicted value can be used to identify the linear relationship between the attributes. Suppose the increase in the product advantage budget will increase the product sales. Based on the problem difference regression algorithms can be used. some of the basic regression algorithms are linear regression, polynomial regression … etc

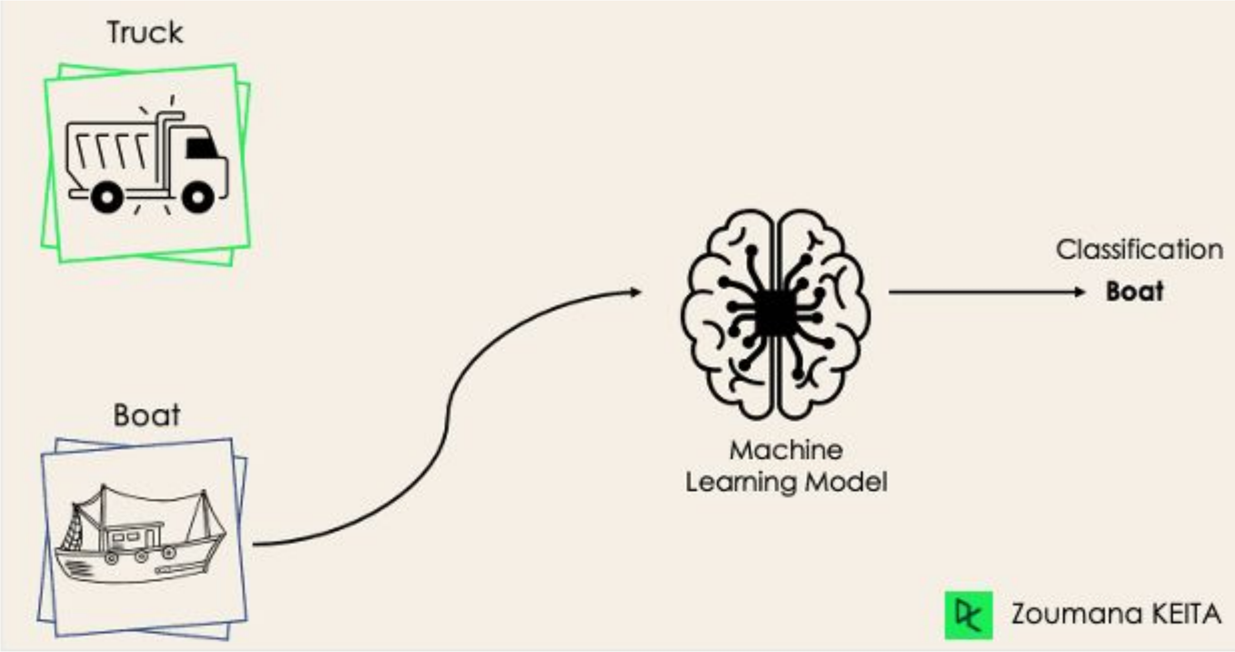
Summary

1. If forecasting target class (Classification).
2. If forecasting a value (Regression).

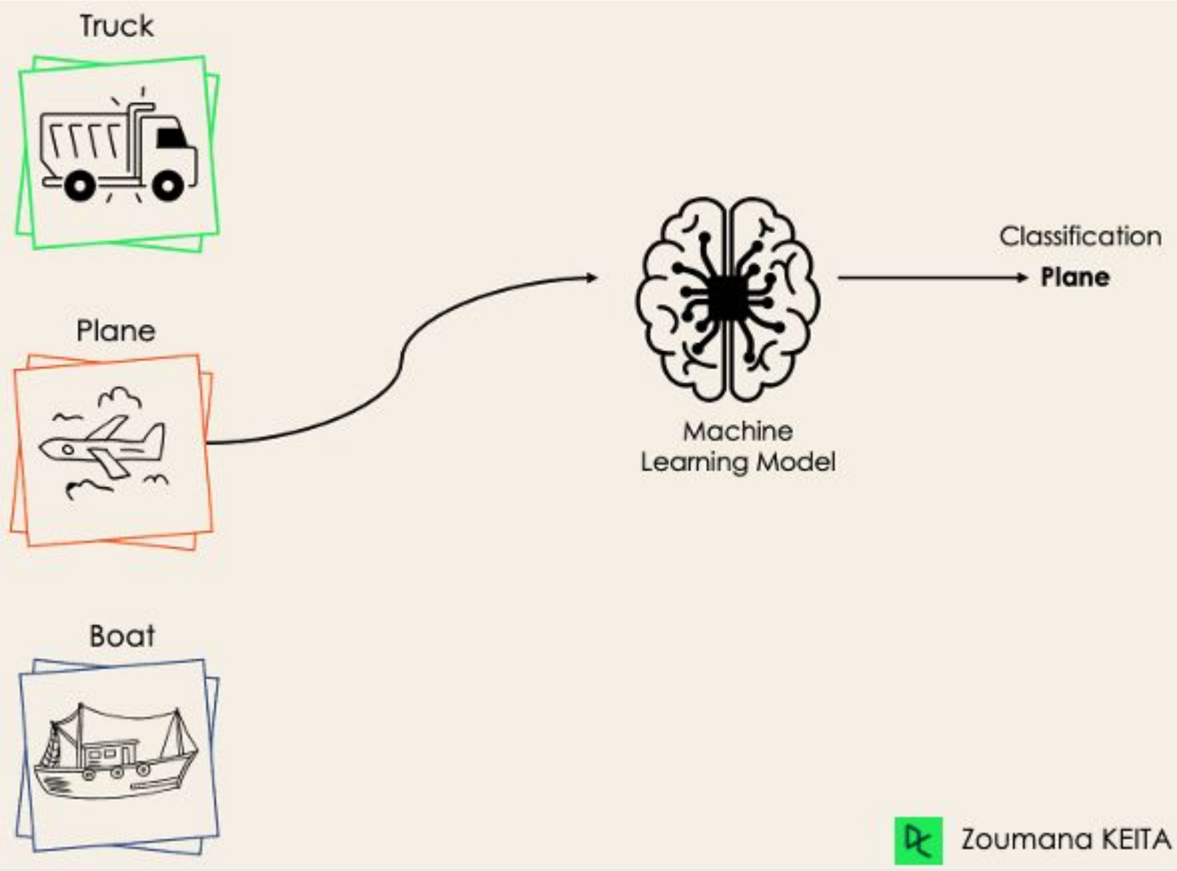


Zoumana KEITA

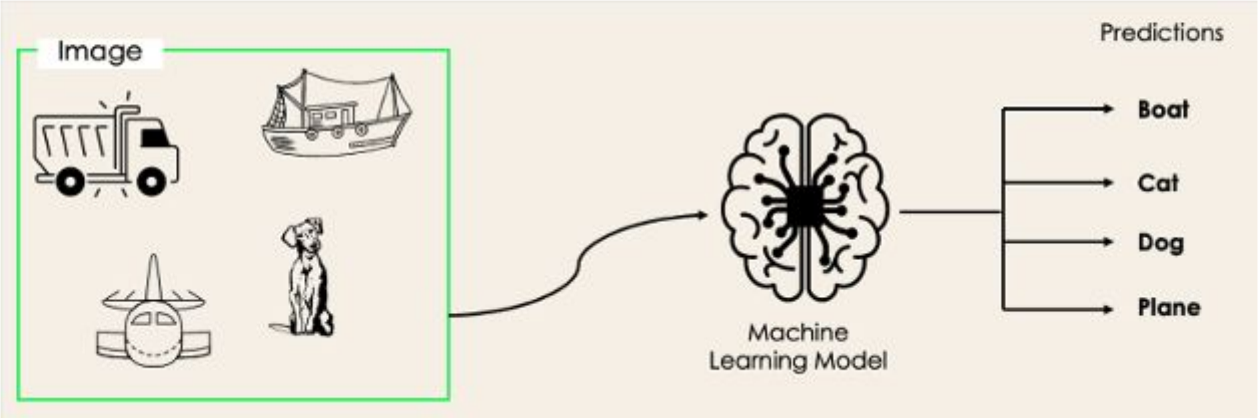
Binary Classification



Multi-Class Classification



Multi-Label Classification



Application of Regression Algorithm

1. **Real Estate Price Prediction:** A regression model can be used to predict the price of a house based on its features, such as size, number of rooms, location, etc.
2. **Income Prediction:** Regression can be used to predict the future income of an individual or a company based on variables like age, education, and work experience.
3. **Time Series Analysis:** Regression is useful for predicting future values in time series data, such as forecasting stock prices, quarterly sales, or product demand over time.
4. **Air Quality Prediction:** Regression models can predict air quality based on variables like pollutant concentration, temperature, and humidity.
5. **Estimating Age of Individuals:** Regression can be used to estimate a person’s age based on facial features, as in facial recognition applications.

Application of Classification Algorithm

1. **Spam Detection**: Spam filters use classification algorithms to determine whether an email is spam or non-spam based on its content and message characteristics.
2. **Medical Diagnosis**: Classification models can assist in medical diagnosis, such as disease detection based on medical test results or medical images.
3. **Document Classification:** Classification is used to organize documents into categories, such as news, emails, reports, etc., in document management applications.
4. **Fraud Detection in Financial Transactions**: Classification algorithms can identify fraudulent transactions based on behavioral patterns and transaction characteristics.
5. **Image Classification:** In computer vision applications, like object recognition, classification is used to label objects in images.

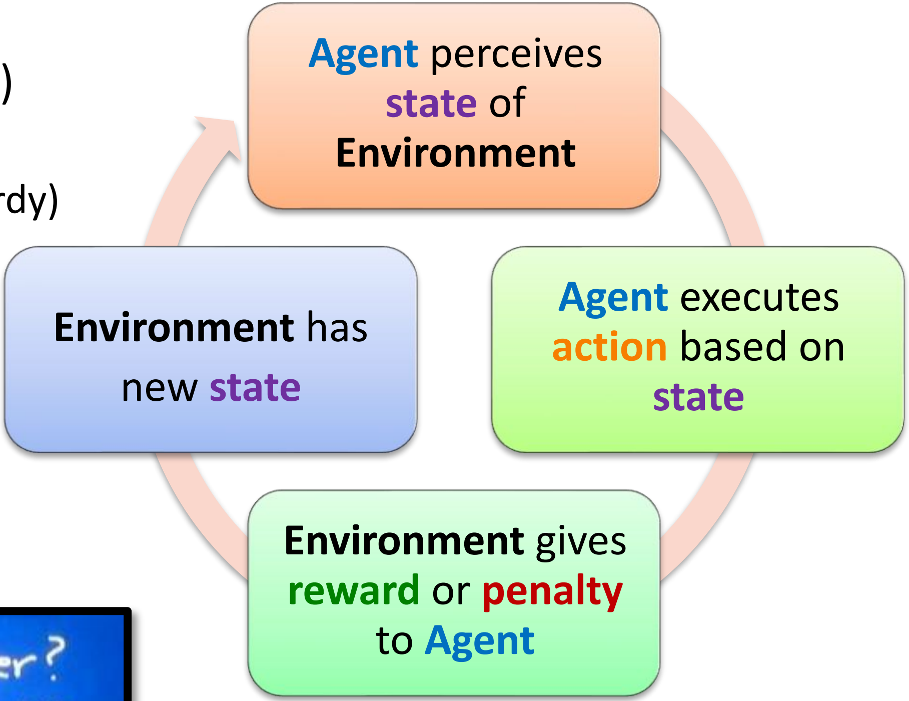
Machine Learning Rudimentary Basics

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



Agent perceives **state** of **Environment**

Agent executes **action** based on **state**

Environment gives **reward** or **penalty** to **Agent**

Environment has new **state**

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

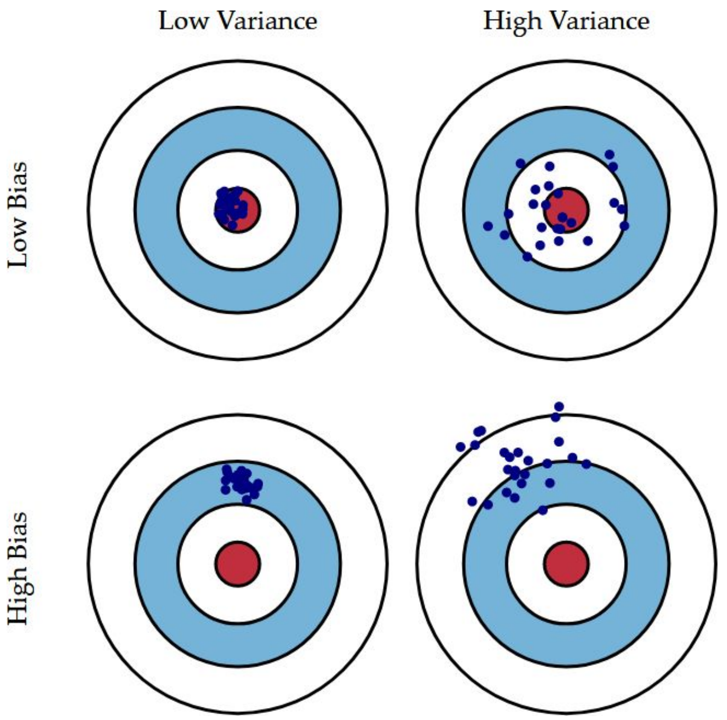
- Measure of inaccuracy
- High bias + low variance consistently inaccurate results

Variance

- Measure of imprecision (lack of reproducibility)
- High variance + low bias inconsistently accurate results

Irreducible error

- Noise that cannot be reduced by optimizing algorithms



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

ML Definitions – Model Evaluation

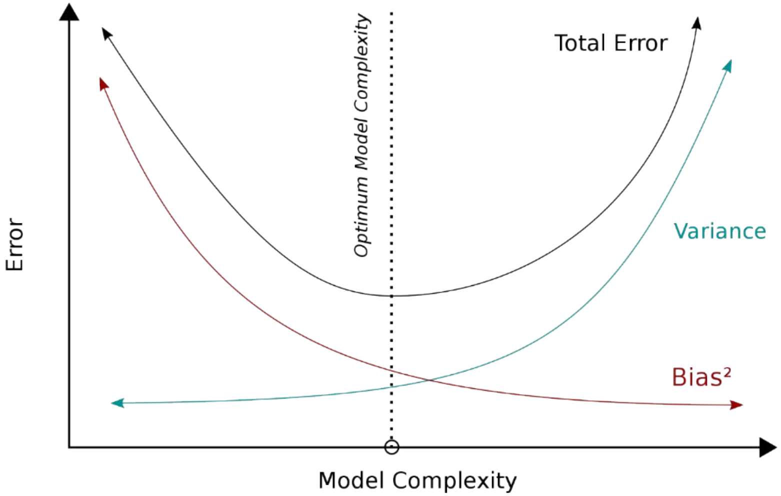
Bias

- *Not just an ethical term…*
- Amount of **<u>inaccuracy</u>** 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 **<u>imprecision</u>** (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

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

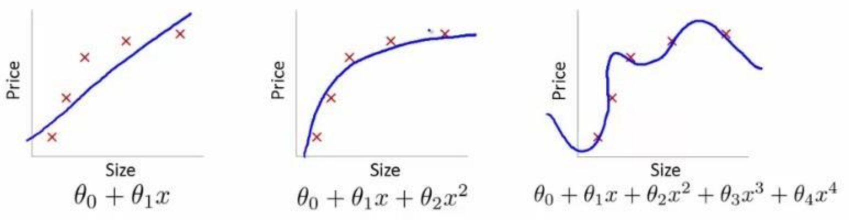
$$error$$

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 <u>exactly</u> fits <u>training</u> 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



High bias (underfit)

"Just right"

High variance (overfit)

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

ML Definitions – Model Evaluation

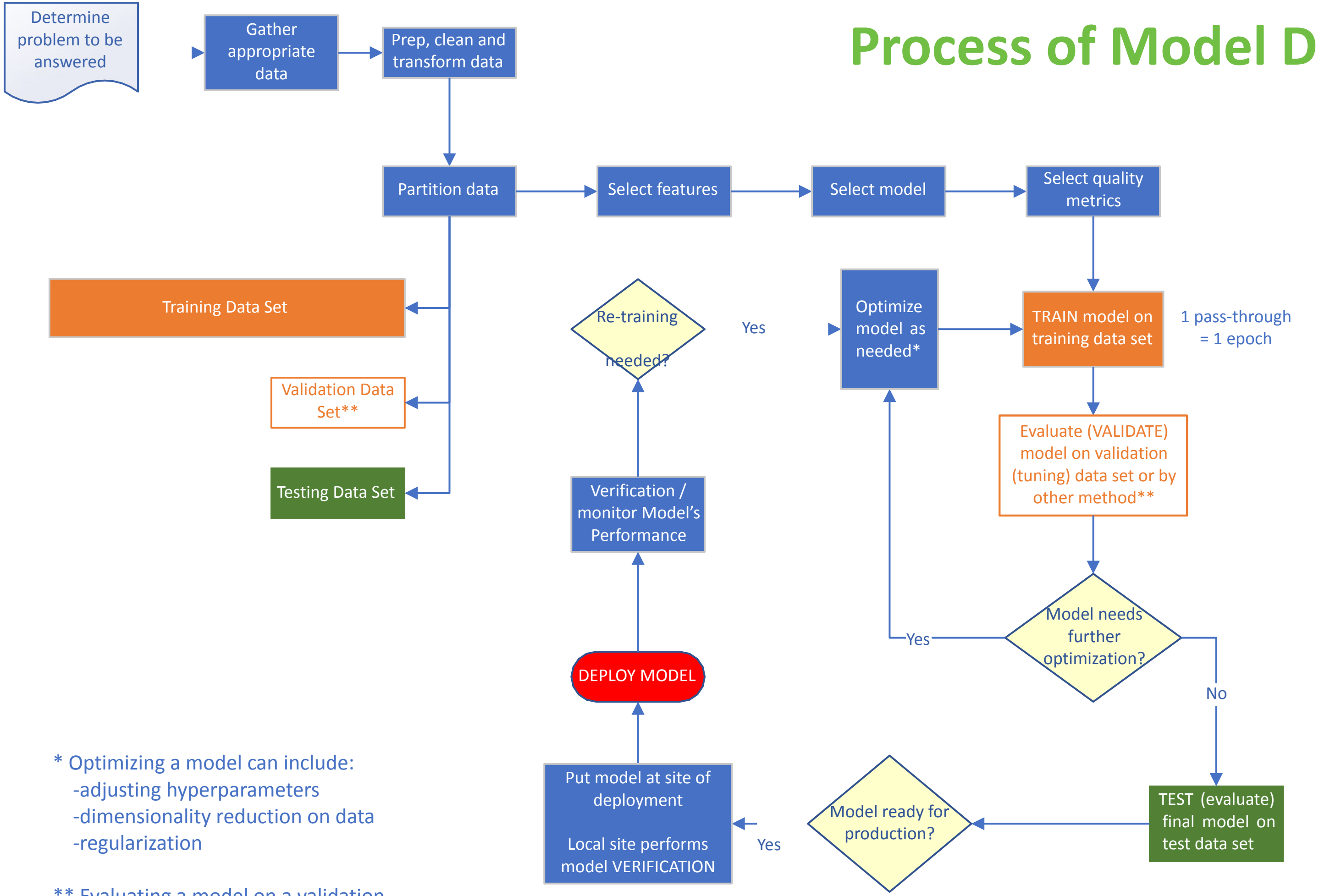
- **Null error rate**
  - For classification methods, rate of being wrong if you ALWAYS pick the majority class
  - If the majority class has 105 instances out of 165 total instances
    - Null error rate = (165 – 105)/165 = 36%
  - **Accuracy paradox**
    - Best classifier for the intended use may 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

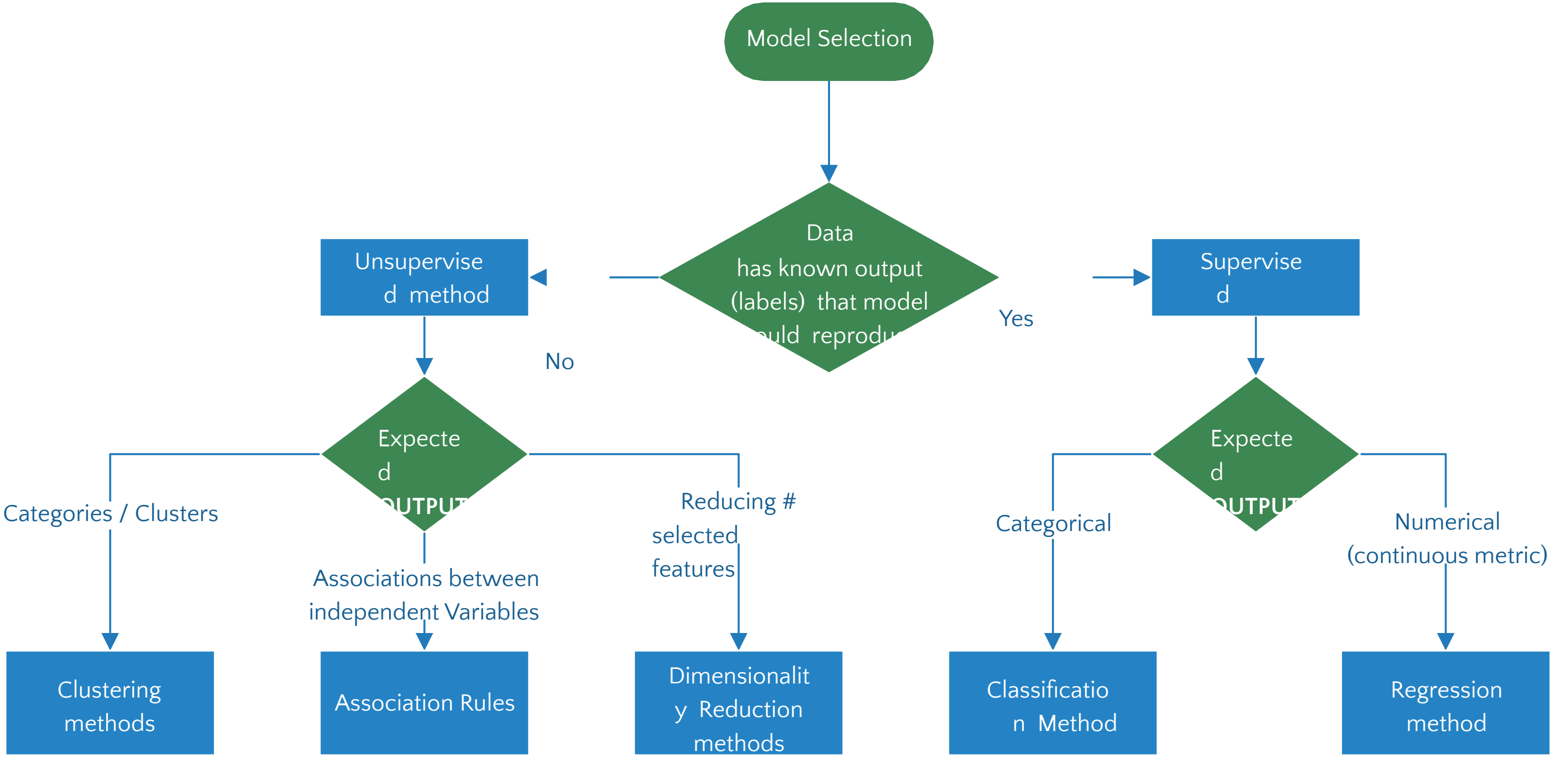
Process of Model Development


Determine problem to be answered
Gather appropriate data
Prep, clean and transform data
Partition data
Select features
Select model
Select quality metrics
Training Data Set
Validation Data Set**
Testing Data Set
Re-training needed?
Yes
Optimize model as needed*
TRAIN model on training data set
1 pass-through = 1 epoch
Evaluate (VALIDATE) model on validation (tuning) data set or by other method**
Model needs further optimization?
Yes
No
TEST (evaluate) final model on test data set
Model ready for production?
Yes
Put model at site of deployment
Local site performs model VERIFICATION
DEPLOY MODEL
Verification / monitor Model's Performance

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

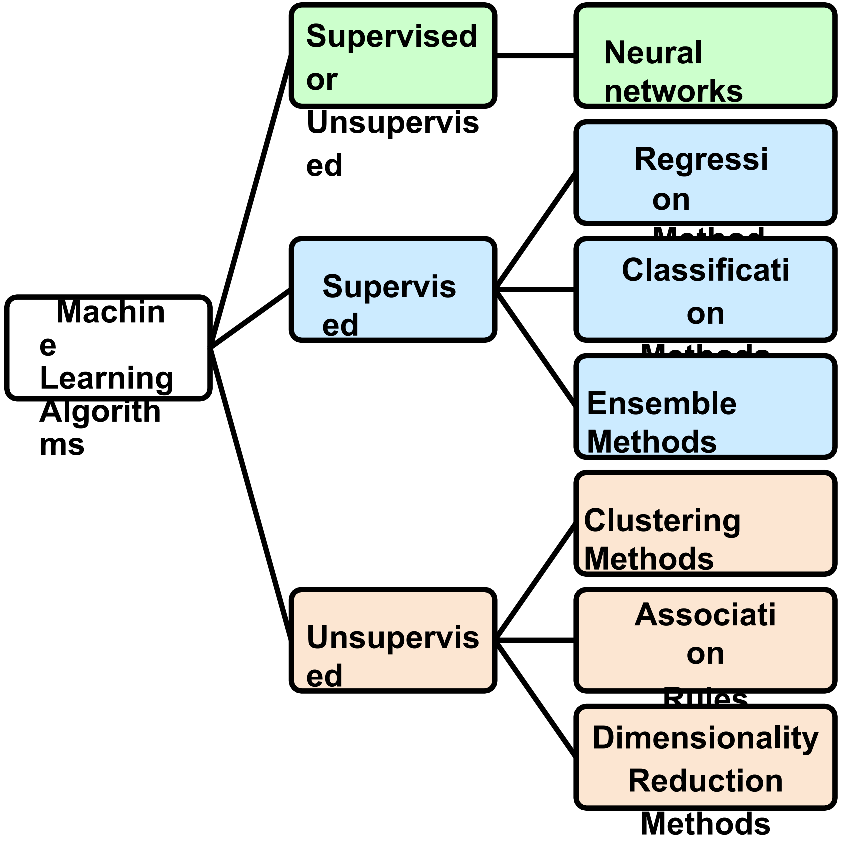
** Evaluating a model on a validation data set may not always be needed.

Process of Model Development



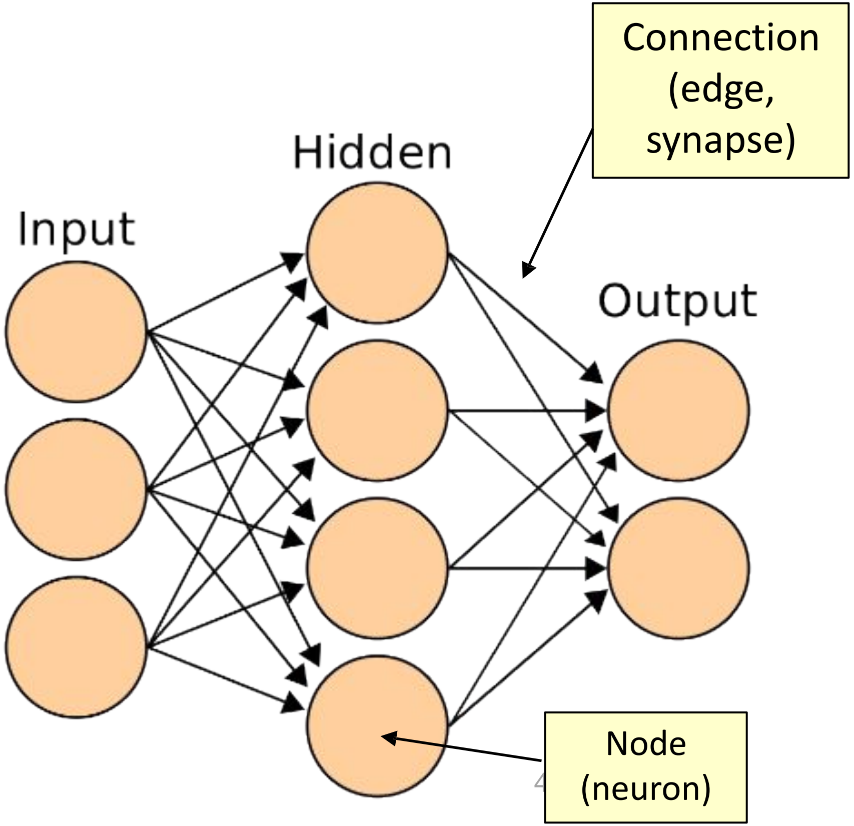
Machine Learning Algorithms

- 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



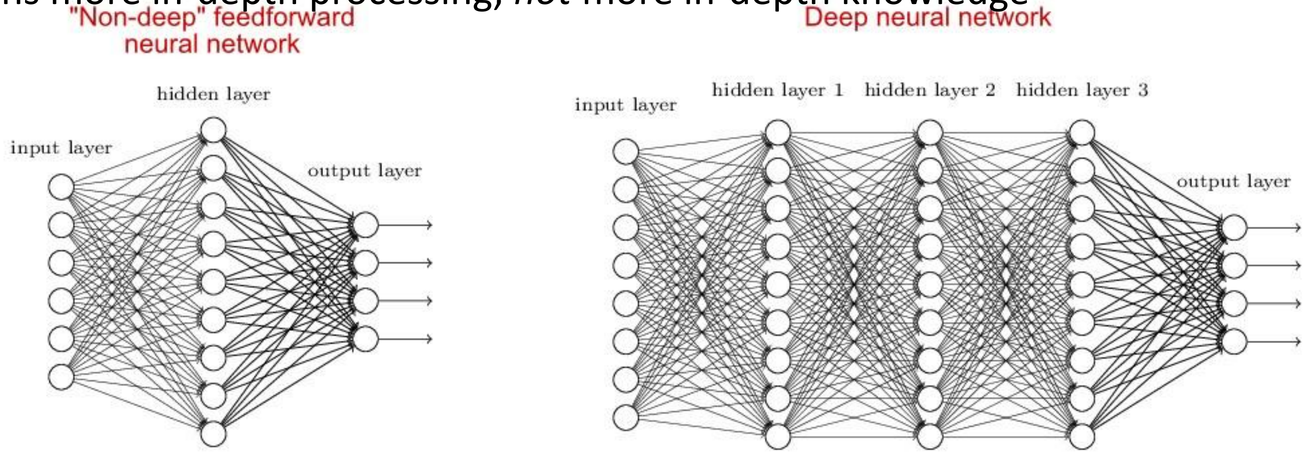
Artificial Neural Networks (ANNs)

- 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?