Applications of AI: NLP, Computer Vision, IoT (UCS655)

Unit – 1

Introduction to Artificial Intelligence

Reference Book

- Artificial Intelligence: A Modern Approach, by Stuart Russell and Peter Norvig, Pearson.
- Artificial Intelligence theory and practice, by T. Dean, J. Allen & Y. Aloimonos, New York: Benjamin Cummings (1995).

Artificial Intelligence

- What is AI
  - Artificial
  - Intelligence
- Building intelligence into man made things
- McCarthy in 1955

Artificial Intelligence (cont.)

- AI things should behave like humans or the idealistic human
- Behavior
  - Thoughts
  - Actions

	Thoughts	**Actions**
Human-like behavior	System that think like humans	System that act like humans
Idealistic behavior	System that think rationally	System that act rationally

Artificial Intelligence (cont.)

- Tasks performed by intelligent entities
  - Mundane
    - Recognizing
    - Communicating
    - Planning any activity
    - Navigating around the obstacles
  - Expert
    - Mathematical problems
    - Medical diagnosis

Components of Intelligent Behavior

- Perception
- Reasoning
- Learning
- Understanding Language
- Solving Problems

AI Systems

- Siri
- Alexa
- Tesla
- Netflix
- Pandora
- Google’s Nest

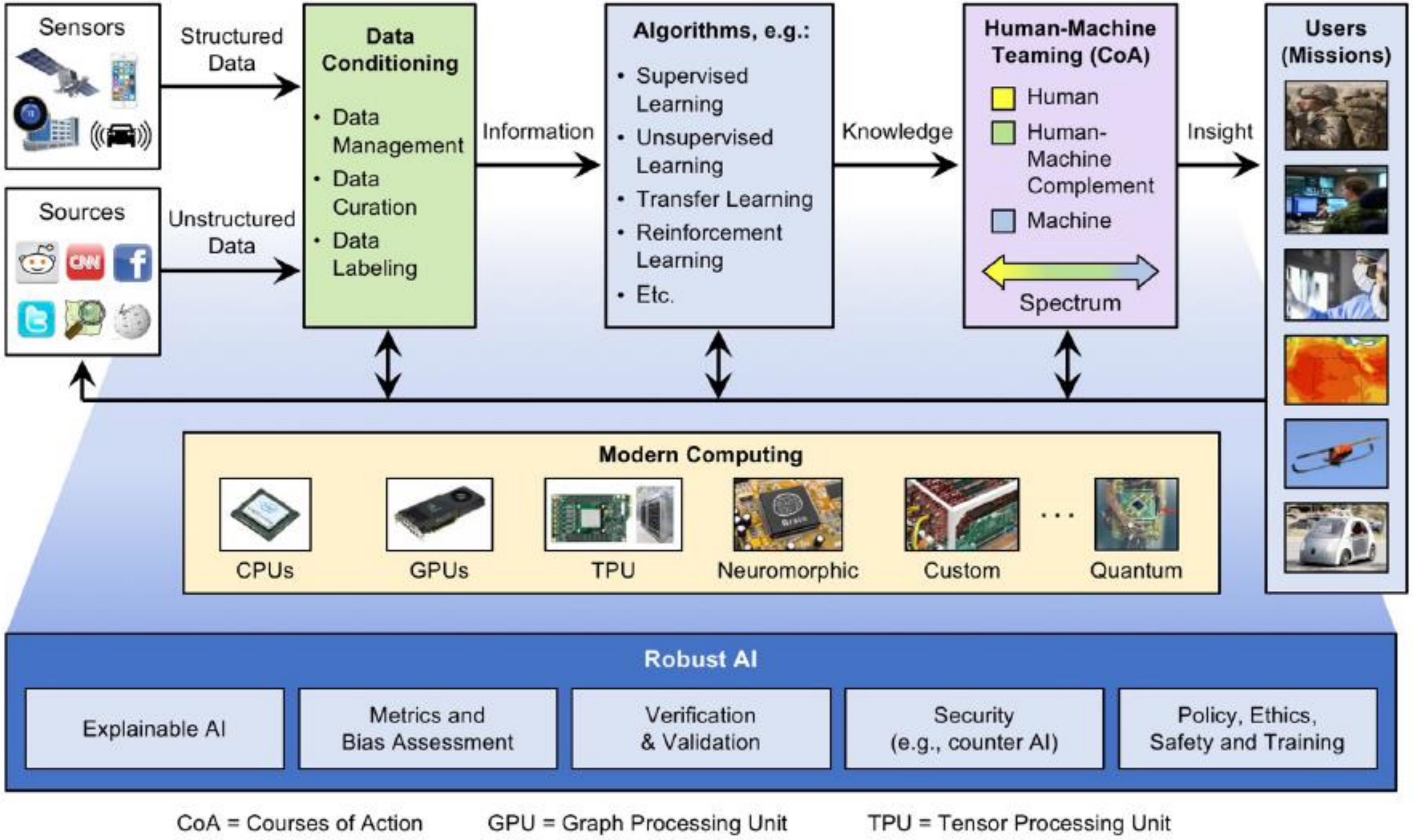
AI: The Umbrella Term

- Machine Learning
  - Deep Learning
- Natural Language Processing
  - Natural Language Understanding (NLU)
  - Natural Language Generation (NLG)
- Computer Vision

Approaches to AI

- Strong AI
- Weak AI
- Applied AI
- Cognitive AI
- Narrow and general AI

AI Architecture



Source: AI Enabling Technologies: A Survey, MIT Lincoln Laboratory

Data Conditioning

- Typically include tasks such as
  - How to store data
  - Handle inconsistencies in the dataset
  - Determining the suitable algorithm
- Data Management
  - Polystore databases: BigDAWG
- Data Curation
  - Anomaly and outlier detection
  - Dimensionality reduction
  - Data weighting and normalization
- Data Labeling

Algorithms

- Supervised Learning
- Unsupervised Learning
- Supervised and Unsupervised neural networks
- Transfer Learning
  - Inductive transfer learning
  - Transductive transfer learning
  - Unsupervised transfer learning
- Semi Supervised Learning
- Reinforcement Learning

Computing

- Processing Technologies
  - CPU
  - GPU
  - TPU
  - Neuromorphic
  - Quantum

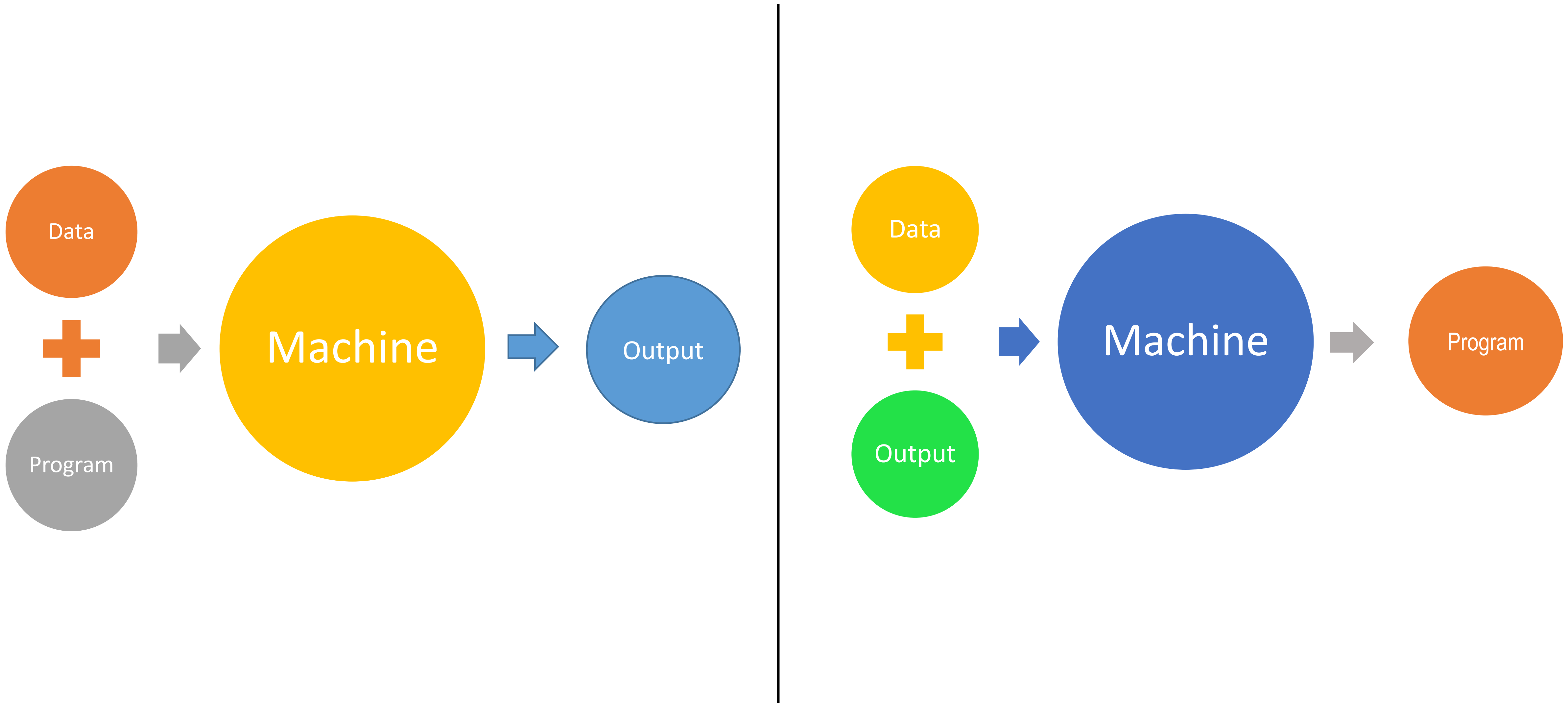
Robust AI

- Explainable AI
- Metrics
- Security

Human Machine Teaming

- Human-in-the loop
- Human-on-the-loop
- Human-out-of-the-loop

Basics of Machine Learning



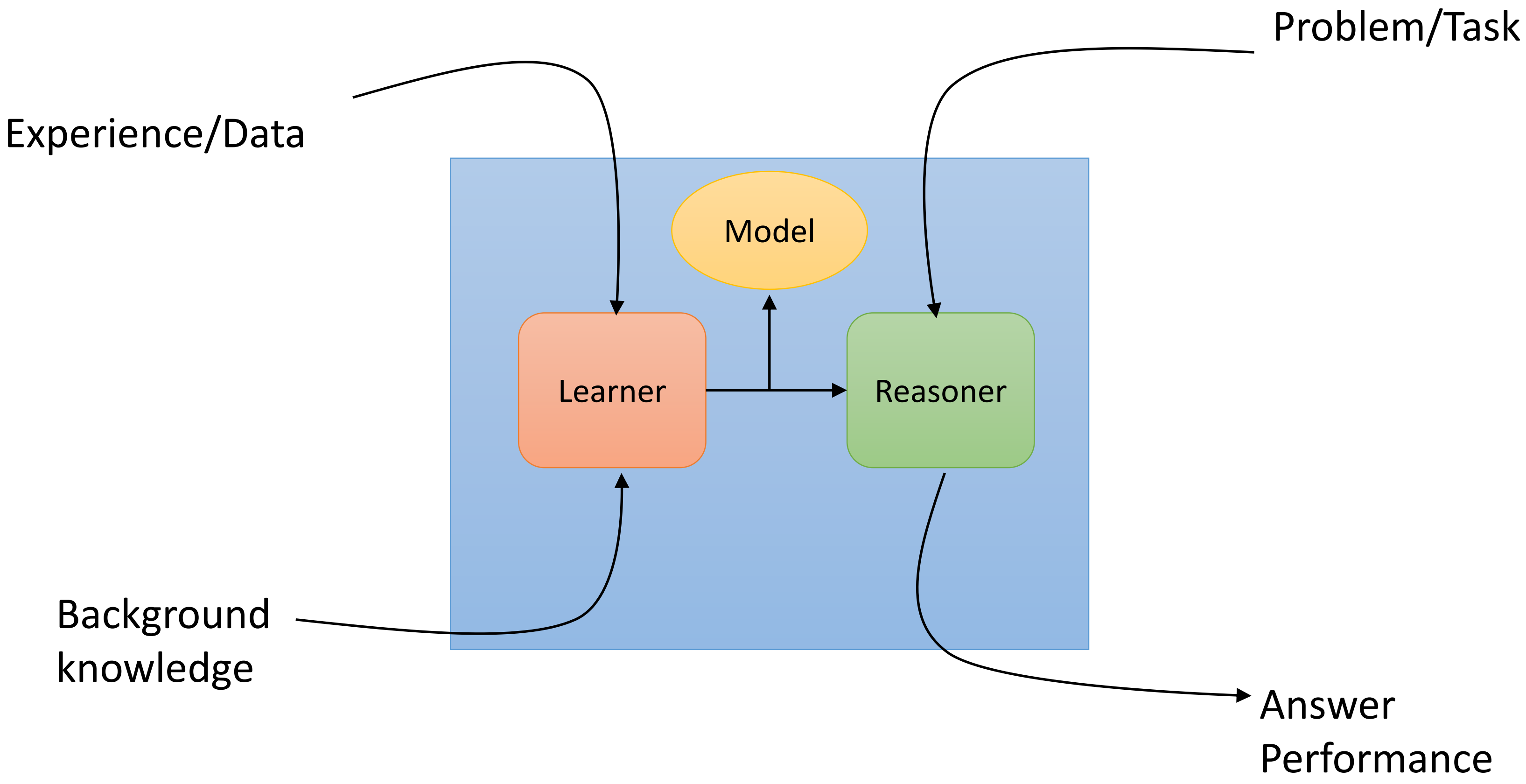
Machine Learning

- Learning
- Machine learning
  - Past data/ Experience – E
  - Class of Tasks – T
  - Performance Measure – P

Components of Learning Algorithm

- Behavior of the task which learning algorithm seeks to improve
- Data or Experience
- Performance measure

Learning System



Machine Learning in Practice

- Medical diagnosis
- Computer vision
- Robot control
- Natural language processing
- Financial domain
- Business intelligence

Steps to create a learner

- Choose the training data
- Choose the target function
- Choose how to represent the target function
- Choose the learning algorithm to infer the target function

Types of Machine Learning

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Semi-supervised Learning

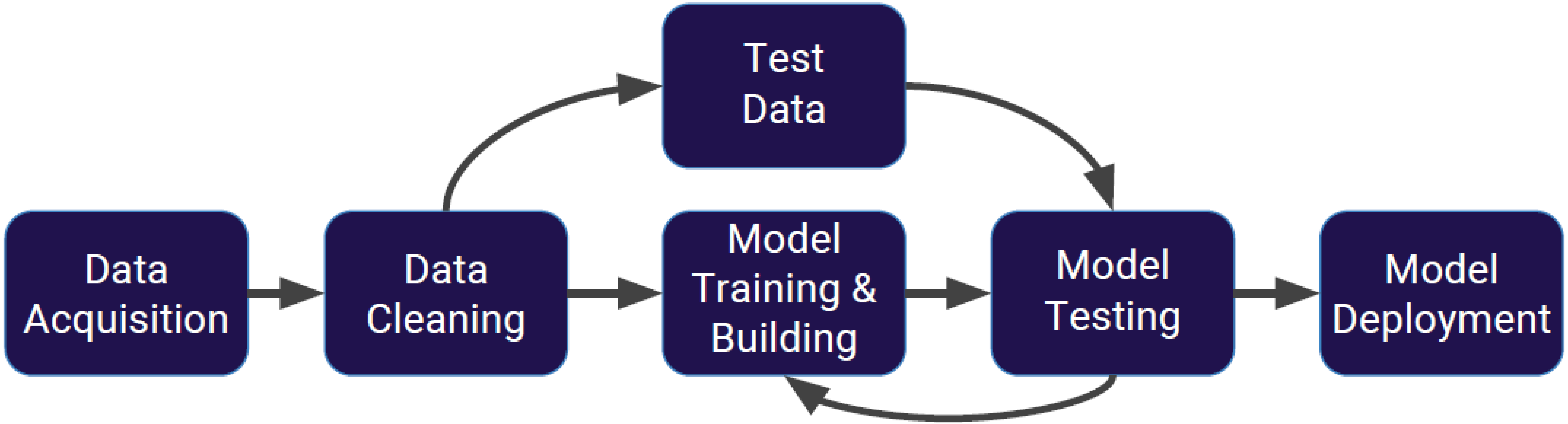
Supervised Learning

- **Supervised learning** algorithms are trained using **labeled** examples, such as an input where the desired output is known.

- For example, a segment of text could have a category label, such as:
  - **Spam** vs. **Legitimate** Email
  - **Positive** vs. **Negative** Movie Review

- Supervised learning is commonly used in applications where historical data predicts likely future events.

Machine Learning Process for Supervised Learning



Cont.

- What we just showed is a simplified approach to supervised learning, it contains an issue!
- Is it fair to use our single split of the data to evaluate our models performance?
- After all, we were given the chance to update the model parameters again and again.

Cont.

- To fix this issue, data is often split into **3 sets**
  - Training Data
    - Used to train model parameters
  - Validation Data
    - Used to determine what model hyperparameters to adjust
  - Test Data
    - Used to get some final performance metric
- This means after we see the results on the **final test set** we don’t get to go back and adjust any model parameters!
- This final measure is what we label the true performance of the model to be.

Supervised Learning

	X_1	X_2	.	.	X_n	Y
I_1	a_1	a_2	.	.	a_n	Y_1
I_2	b_1	b_2	.	.	b_n	Y_2
	.				.	
	.				.	
	.				.	
I_n	z_1	z_2	.	.	z_n	Y_n
Test Instance	t_1	t_2	.	.	t_n $\rightarrow$	?

Classification Problem Example

- Credit Scoring
  - Income
  - Savings

Regression Problem Example

- Predict price of a car
  - Mileage of the car

Feature and its types

- Categorical
- Ordinal
- Integer
- Real-valued

Example

Training Examples:

	Action	Author	Thread	Length	Where
e1	skips	known	new	long	Home
e2	reads	unknown	new	short	Work
e3	skips	unknown	old	long	Work
e4	skips	known	old	long	home
e5	reads	known	new	short	home
e6	skips	known	old	long	work

New Examples:

e7	???	known	new	short	work
e8	???	unknown	new	short	work

Components of Classification Learning

- Task, T
  - Task: Medical diagnosis
  - Instance: patient health record
- Performance metric, P
- Experience, E

Representation of the function

- Linear function
- Decision Tree
- Multivariate linear function
- Single layer perceptron
- Multi-layer neural network

Basic Terminologies

- Feature
- Instance Space/ Feature space (X)
- Training data
- Concept (c)
- Target function (f)
- Hypotheses (h)
- Hypotheses Space (H)

Feature space and Hypotheses space calculation

- Boolean feature: 4
- Number of possible instances
- Number of possible Boolean functions

Inductive Bias

- Restriction Bias
- Preference Bias

Inductive Learning

- Given training examples → General function
- Machine learning is all about generalization
  - Bias error
  - Variance error
- Overfitting and Underfitting

Evaluation of Learning Algorithm

- Experimental evaluation
  - Error
  - Accuracy
  - Precision
  - Recall

Train-Test Split

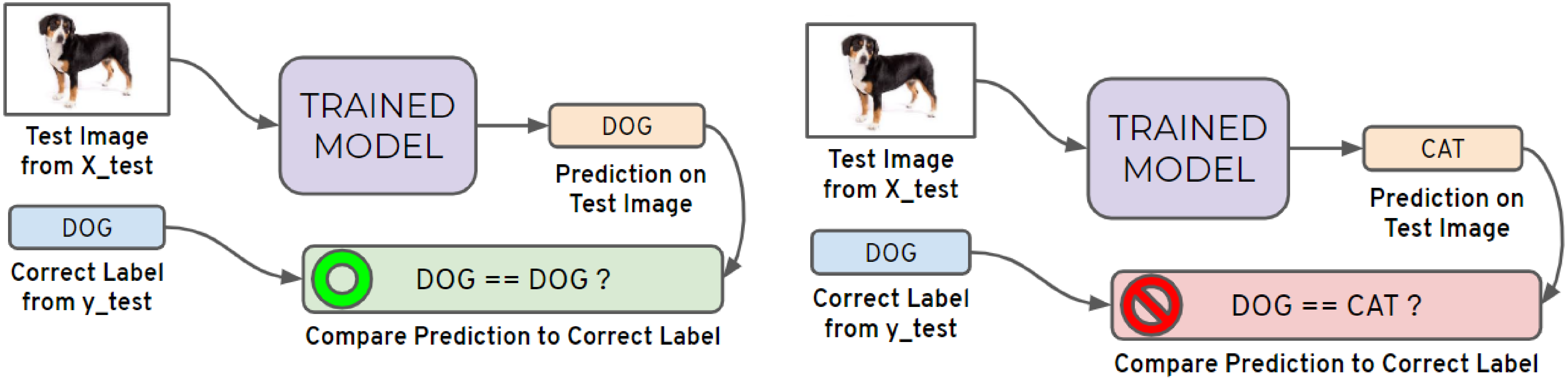
- While you could get these measurements using the same data you trained your model on, that is not a good idea.
- Your model has already seen this data meaning it is not a good choice for evaluating your model's performance
- You should get these metrics off test data, which your model has not seen yet.
- This is known as a train-test split.

Model Evaluation

- The key classification metrics we need to understand are:
  - Accuracy
  - Recall
  - Precision
  - F1-Score
- Typically in any classification task your model can only achieve two results:
  - Either your model was **correct** in its prediction.
  - Or your model was **incorrect** in its prediction.

Example

- we will attempt to predict if an image is a dog or a cat.
- Once we have the model's predictions from the **X_test** data, we compare it to the **true y values** (the correct labels).



Accuracy

- Accuracy
  - Accuracy in classification problems is the **number of correct predictions** made by the model divided by the **total number of prediction**
- For example, if the X_test set was 100 images and our model **correctly** predicted 80 images, then we have **80/100.**
  - **0.8** or **80% accuracy.**
- Accuracy is useful when target classes are well balanced
  - we would have roughly the same amount of cat images as we have dog images.
- Downside
  - Accuracy is **not** a good choice with **unbalanced** classes!
  - Imagine we had 99 images of dogs and 1 image of a cat.
  - If our model was simply a line that always predicted **dog** we would get 99% accuracy!

Recall and Precision

- Recall
  - Ability of a model to find all the relevant cases within a dataset.
  - The precise definition of recall is the number of true positives **divided by** the number of true positives plus the number of false negatives.
- Precision
  - Ability of a classification model to identify only the relevant data points.
  - Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.
- Recall and Precision
  - Often you have a trade-off between Recall and Precision.
  - While recall expresses the ability to find all relevant instances in a dataset, precision expresses the proportion of the data points our model says was relevant actually were relevant

F1 Score

- F1-Score
  - In cases where we want to find an optimal blend of precision and recall we can combine the two metrics using what is called the F1 score.
  - The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation:
  - $F1 = 2 * \frac{precision * recall}{precision + recall}$
- We use the harmonic mean instead of a simple average because it punishes extreme values.
- A classifier with a precision of 1.0 and a recall of 0.0 has a simple average of 0.5 but an F1 score of 0.

Confusion Matrix

		predicted condition	
	total population	prediction positive	prediction negative
true condition	condition positive	**True Positive (TP)**	**False Negative (FN)** (type II error)
	condition negative	**False Positive (FP)** (Type I error)	**True Negative (TN)**

		predicted condition		
	total population	prediction positive	prediction negative	Prevalence $= \frac{\Sigma \text{ condition positive}}{\Sigma \text{ total population}}$
true condition	condition positive	**True Positive (TP)**	**False Negative (FN)** (type II error)	True Positive Rate (TPR), Sensitivity, Recall, Probability of Detection $= \frac{\Sigma \text{ TP}}{\Sigma \text{ condition positive}}$
	condition negative	**False Positive (FP)** (Type I error)	**True Negative (TN)**	False Positive Rate (FPR), Fall-out, Probability of False Alarm $= \frac{\Sigma \text{ FP}}{\Sigma \text{ condition negative}}$
	Accuracy $= \frac{\Sigma \text{ TP} + \Sigma \text{ TN}}{\Sigma \text{ total population}}$	Positive Predictive Value (PPV), Precision $= \frac{\Sigma \text{ TP}}{\Sigma \text{ prediction positive}}$	False Omission Rate (FOR) $= \frac{\Sigma \text{ FN}}{\Sigma \text{ prediction negative}}$	Positive Likelihood Ratio (LR+) $= \frac{\text{TPR}}{\text{FPR}}$
		False Discovery Rate (FDR) $= \frac{\Sigma \text{ FP}}{\Sigma \text{ prediction positive}}$	Negative Predictive Value (NPV) $= \frac{\Sigma \text{ TN}}{\Sigma \text{ prediction negative}}$	Negative Likelihood Ratio (LR–) $= \frac{\text{FNR}}{\text{TNR}}$

Cont.

- The main point to remember with the confusion matrix and the various calculated metrics is that they are all fundamentally ways of comparing the predicted values versus the true values.
- What constitutes “good” metrics, will really depend on the specific situation!
- What is a good enough accuracy?
  - Did you create a model to predict presence of a disease?
  - Is the disease presence well balanced in the general population? (Probably not!)
- Often we have a precision/recall trade off, We need to decide if the model will should focus on fixing False Positives vs. False Negatives.
  - In disease diagnosis, it is probably better to go in the direction of False positives, so we make sure we correctly classify as many cases of disease as possible!

Regression Evaluation

- Regression is a task when a model attempts to predict continuous values (unlike categorical values, which is classification)
- The evaluation metrics like accuracy or recall aren’t useful for regression problems, we need metrics designed for **continuous** values!
- For example, attempting to predict the price of a house given its features is a **regression task.**
- Attempting to predict the country a house is in given its features would be a classification task.

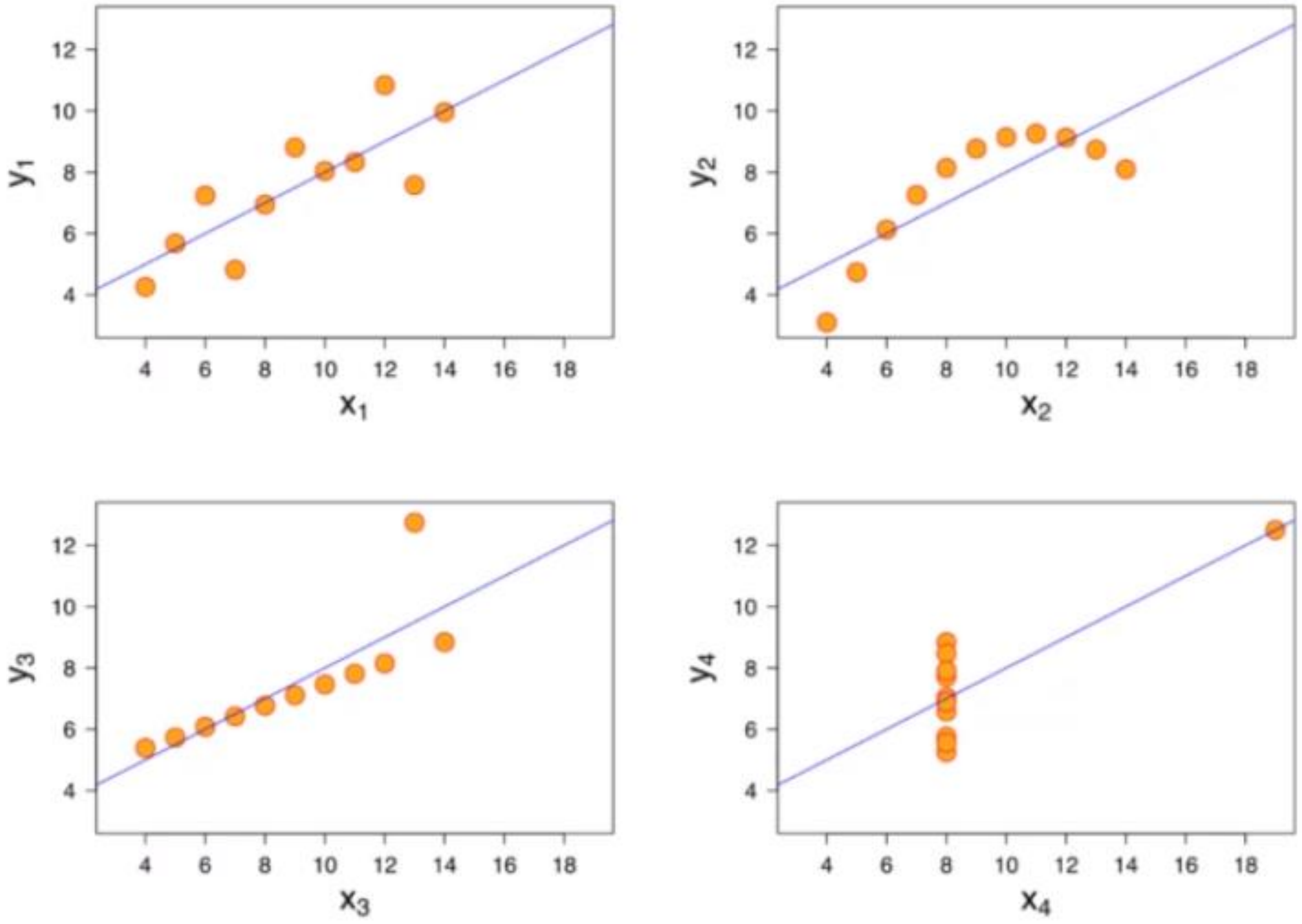
Cont.

- The most common evaluation metrics for regression:
  - Mean Absolute Error
  - Mean Squared Error
  - Root Mean Square Error
- Mean Absolute Error (MAE)
  - This is the mean of the absolute value of errors.
  - Easy to understand

$$= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

Cont.

- MAE won’t punish large errors.



Cont.

- Mean Squared Error (MSE)
  - This is the mean of the squared errors.
  - Larger errors are noted more than with MAE, making MSE more popular.

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- Root Mean Square Error (RMSE)
  - This is the root of the mean of the squared errors.
  - Most popular (has same units as y)

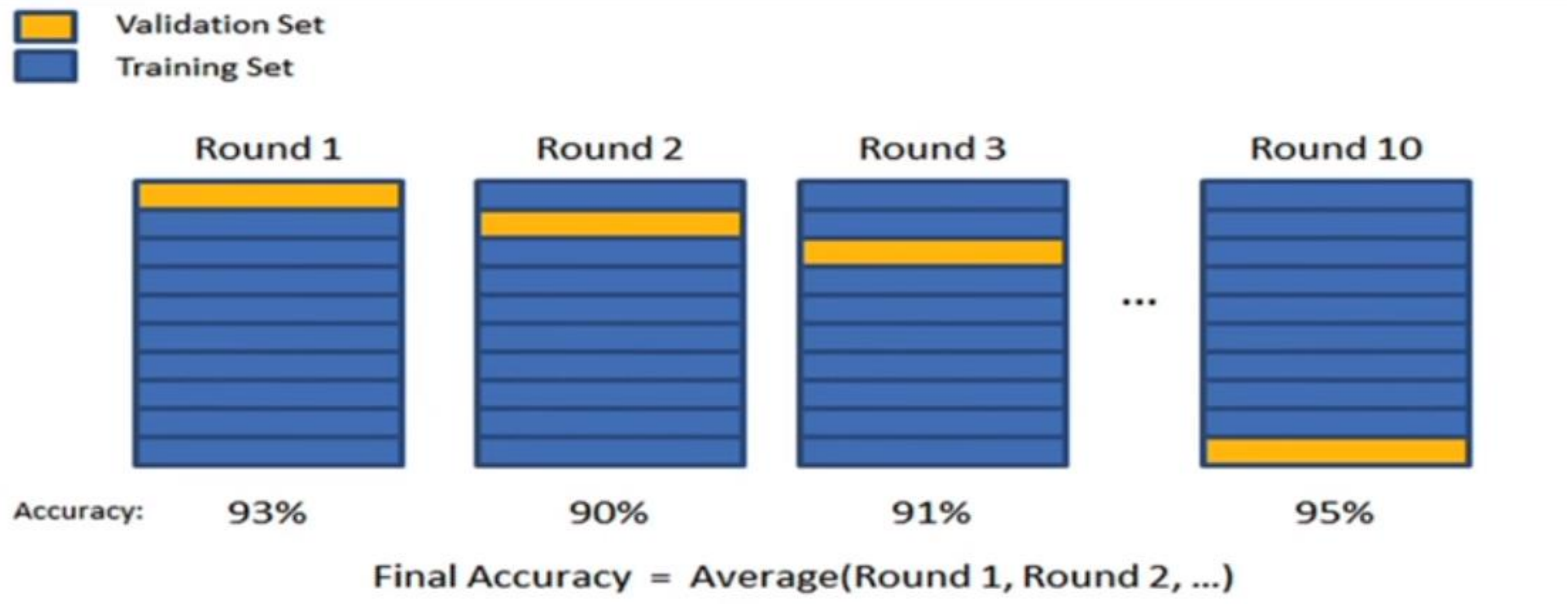
$$\sqrt{= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Cont.

- Is this value of RMSE good?
    - Context is everything!
- A RMSE of $10 is fantastic for predicting the price of a house, but horrible for predicting the price of a candy bar!

Cross-Validation

- K-fold cross validation



Model Selection for Machine Learning

- Model Selection
- Model selection vs assessment
- Model Selection Approaches
  - Probabilistic Measures
  - Resampling Methods

Probabilistic Measures

- Akaike Information Criterion (AIC)
  - $AIC \ = \ 2k - 2\ln(\hat{L})$
  - Model with lowest AIC score is selected
  - Relative likelihood = $e^{(AIC_1 - AIC_2)/2}$
- Bayesian Information Criterion (BIC)
  - $BIC \ = \ kln(n) - 2\ln(\hat{L})$
  - Model with lowest BIC score is selected
  - Probability that BIC will select correct model approaches 1 as n approaches ∞
- Minimum Description Length (MDL)
  - MDL = $L(h) - L(D|h)$
  - Model with lowest MDL is selected

Resampling Methods

- Random train/test splits
- Cross-validation
- Bootstrap

Bootstrap Method

- Choose a number of bootstrap samples to perform
- Choose a sample size
- For each bootstrap sample
  - Draw a sample with replacement with the chosen size
  - Fit a model on the sample data
  - Estimate the skill of the model on out-of-bag (OOB) sample
- Calculate the mean of the sample of model skill estimates

Five Tribes of Machine Learning

- Pedro Domingos – The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World
- Symbolists
    - Root – Logic
    - Master algorithm – Inverse deduction
- Connectionists
    - Root – Neuroscience
    - Master algorithm - Backpropagation
- Bayesians
    - Root – Statistics
    - Master algorithm – Probabilistic Inference using the Bayes Theorem
- Evolutionaries
    - Root – Evolutionary Biology
    - Master algorithm – Genetic Programming
- Analogizers
    - Root – Psychology
    - Master algorithm – Support vector machine

Symbolists

- Tom Mitchell
- Steve Muggleton
- Ross Quinlan
- Induction : specific facts to general rules

2 + 2 = ?

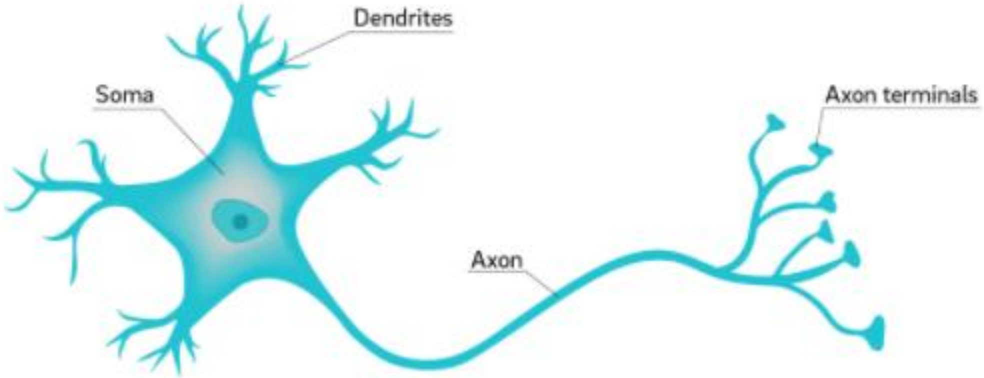
2 + ? = 4

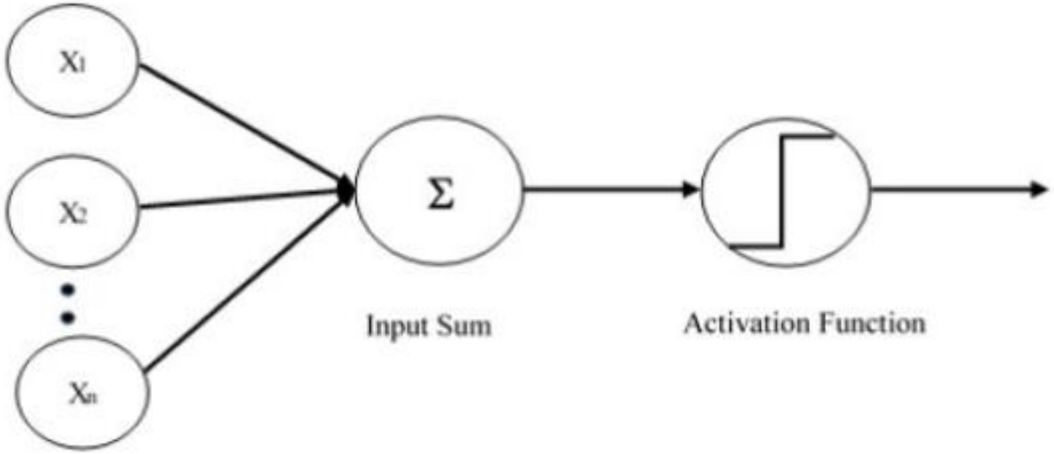
X is human + Humans are mortal = ?

X is human + ? = X is mortal

Connectionists

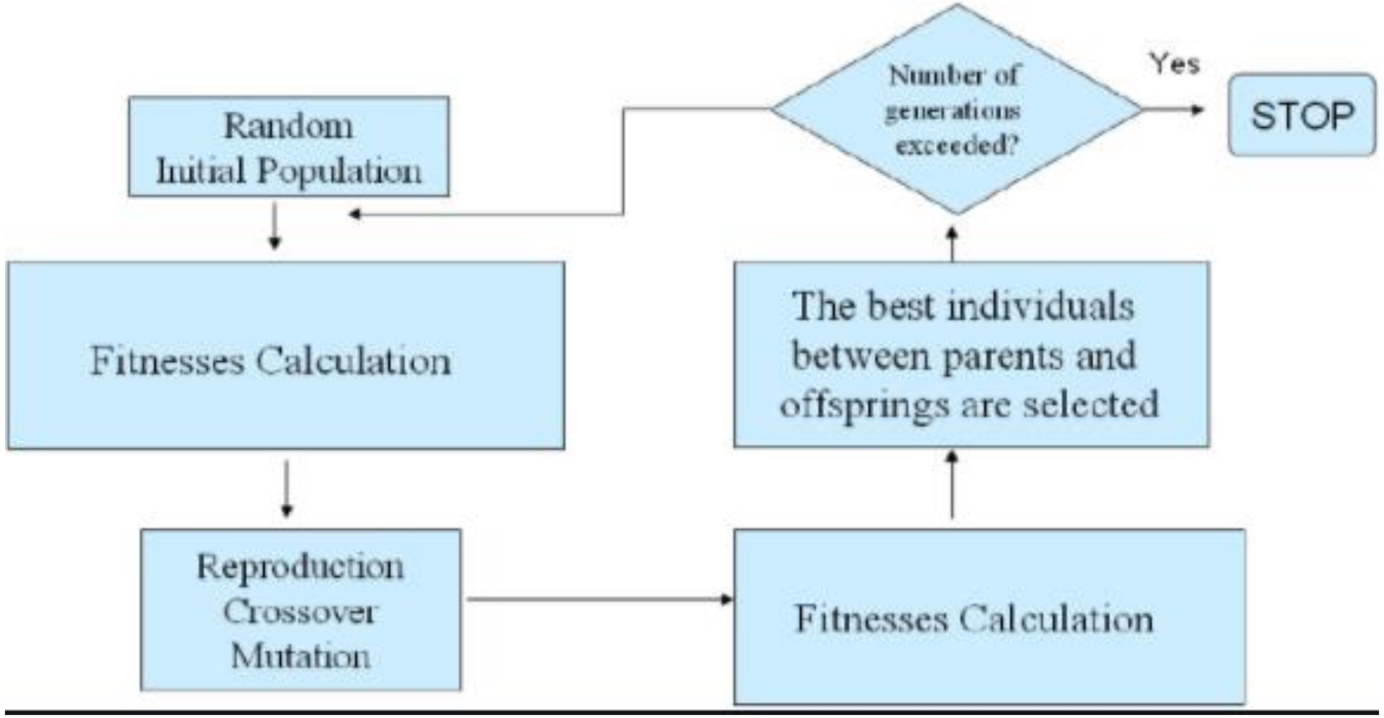
- Geoff Hinton
- Yann LeCun
- Yoshua Bengio





Evolutionaries

- John Holland
- John Koza
- Hod Lipson
- Create a program that mimics evolution



Bayesians

- Judea Pearl – Bayesian network, Turing Award
- David Heckerman
- Michael Jordan
- Probabilistic Inference
- $P(H|e) = \frac{P(e|H)P(H)}{P(e)}$
  - $P(H|e) - \ Posterior$
  - $P(H) - Prior$
  - $P(e|H) - Likelihood$
  - $P(e) - Marginal$

Analogizers

- Reason by similarity
- Peter Hart – nearest neighbor algorithm
- Vladimir Vapnik – Support vector machine
- Douglas Hofstadter
- Most common example – Recommender system