MACHINE LEARNING

REFERENCE BOOKS

- Machine Learning in Action by Peter Harrington publisher Manning Shelter Island
- Understanding Machine Learning From Theory to Algorithms By Shai Shalev-Shwartz and Shai Ben David Cambridge University Press
- MaCHINE Learning by Tom Mitchell
- Introduciton to machine Learning Etham Alpaydin MIT Press

Outline

- Understand Machine Learning
- Supervised Learning- Classification
- Supervised Learning- Regression
- Support Vector Machine
- Improving Classification- AdaBoost
- Unsupervised Learning
- Additional Core Components – PCA

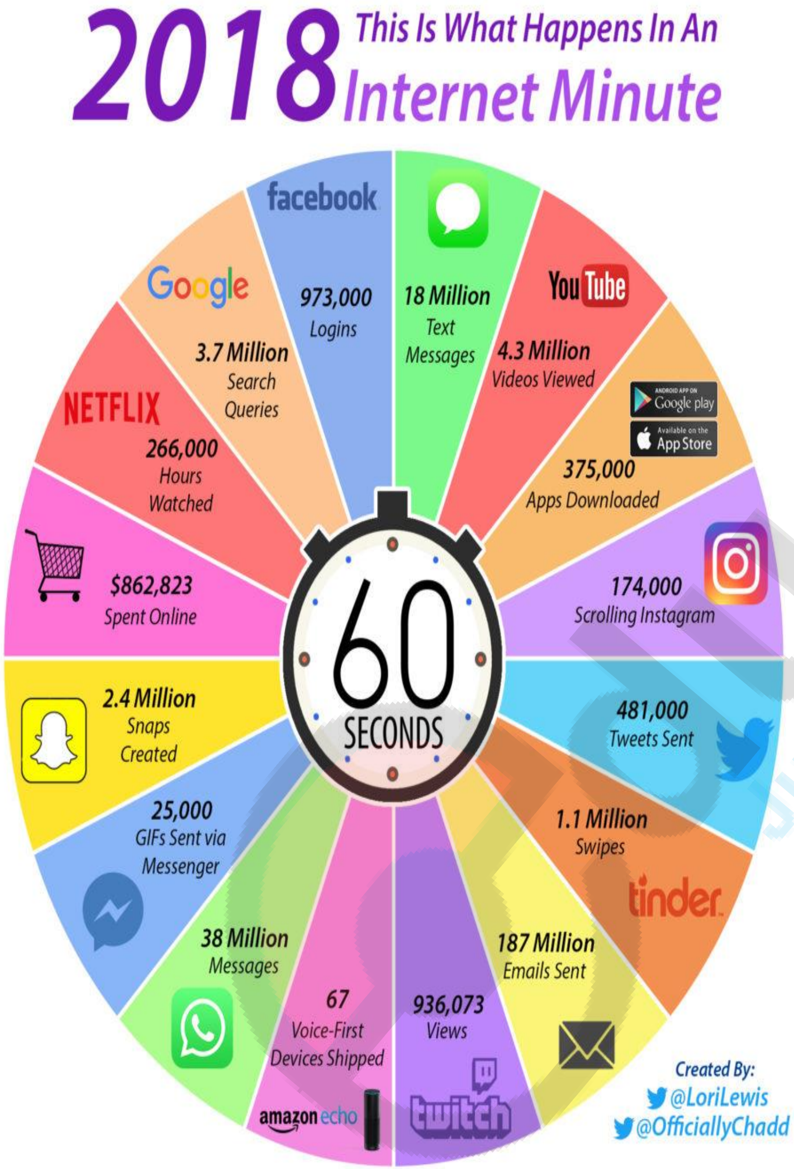
Introduction

- Overview of Machine Learning,
- Key Terminology and task of ML,
- Applications of ML,
- Software Tools

- Machine learning is actively being used today, perhaps in many more places than you'd expect.

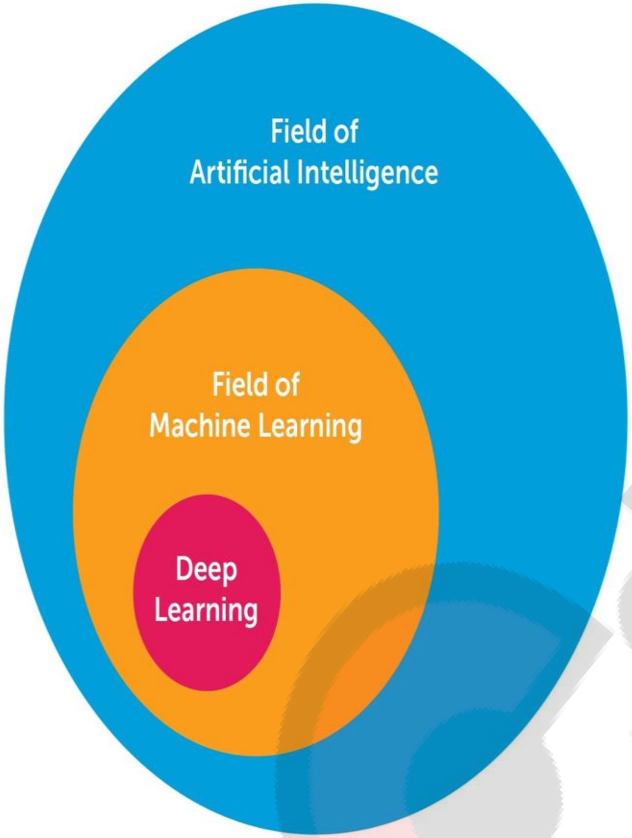

Today's Recommendations F
Here's a daily sample of items recomm
for you. Click here to see all recommendations.
The Manga Guide to the Univers... (Paperback) by Kenji
7 more
Starred
Chats
All Mail
Spam (106)
Trash
Receipts
Work
Manage labels

DATA IS THE NEW OIL



Yesterday, more data was created on our planet than the entire data created in the 20th century

WHAT IS DATA SCIENCE



Data Science

is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data.

What is AI

The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages

AI

ML

DL

NLP

Machine learning

- Insight or knowledge you're trying to get out of the raw data won't be obvious from looking at the data.
- For example, in detecting spam email, looking for the occurrence of a single word may not be very helpful.
- But looking at the occurrence of certain words used together, combined with the length of the email and other factors, you could get a much clearer picture of whether the email is spam or not.
- Machine learning is turning data into information.

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience.
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases

Why “Learn” ?

- Machine learning lies at the intersection of computer science, engineering, and statistics and often appears in other disciplines.
- Machine learning uses statistics.
- we don’t know enough about the problem or don’t have enough computing power to properly model the problem.
- For these problems we need statistics.
- For example, the motivation of humans is a problem that is currently too difficult to model.

Machine Learning?

- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science: Efficient algorithms to
  - Solve the optimization problem
  - Representing and evaluating the model for inference

What We Talk About When We Talk About "Learning"

- Learning general models from a data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior:

  People who bought "Blink" also bought "Outliers" (www.amazon.com)
- Build a model that is *a good and useful approximation* to the data.

- ***Sensors and the data deluge***
- ***Machine learning will be more important in the future***
  - world has moved from manual labor to what is known as *knowledge work*.
- With so much of the economic activity dependent on information, you can't afford to be lost in the data.
- Machine learning will help you get through all the data and extract some information

Key terminology

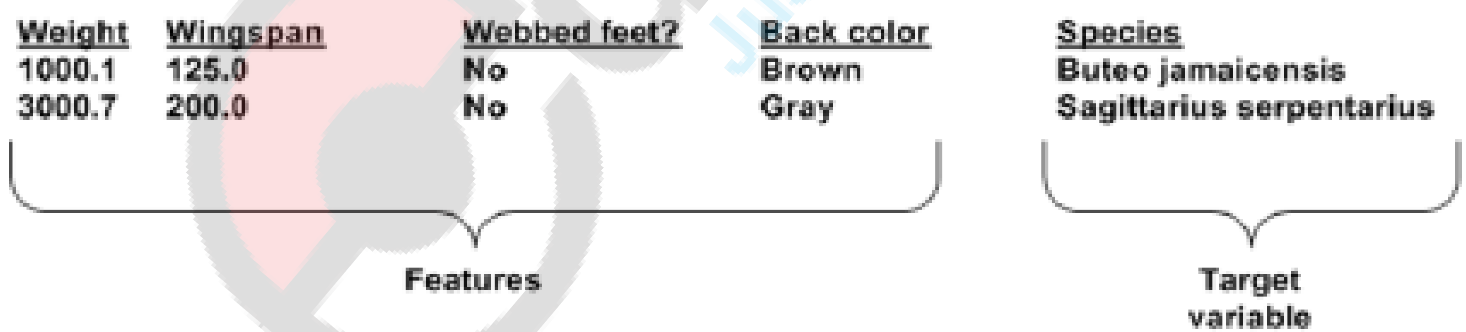
Table 1.1 Bird species classification based on four features

	Weight (g)	Wingspan (cm)	Webbed feet?	Back color	Species
1	1000.1	125.0	No	Brown	Buteo jamaicensis
2	3000.7	200.0	No	Gray	Sagittarius serpentarius
3	3300.0	220.3	No	Gray	Sagittarius serpentarius
4	4100.0	136.0	Yes	Black	Gavia immer
5	3.0	11.0	No	Green	Calothorax lucifer
6	570.0	75.0	No	Black	Campephilus principalis

- Features or attributes are the individual measurements that, when combined with other features, make up a training example.
- This is usually columns in a training or test set.

Learning Method

- To test machine learning algorithms what's usually done is to have a
  - training set data
  - *test set*.
- **Step 1** : the program is fed the training examples; this is when the machine learning takes place.
- **Step 2** : test set is fed to the program.
- The target variable for each example from the test set isn't given to the program, and the program decides which class each example should belong to.
- The target variable or class that the training example belongs to is then compared to the predicted value, and we can get a sense for how accurate the algorithm is.

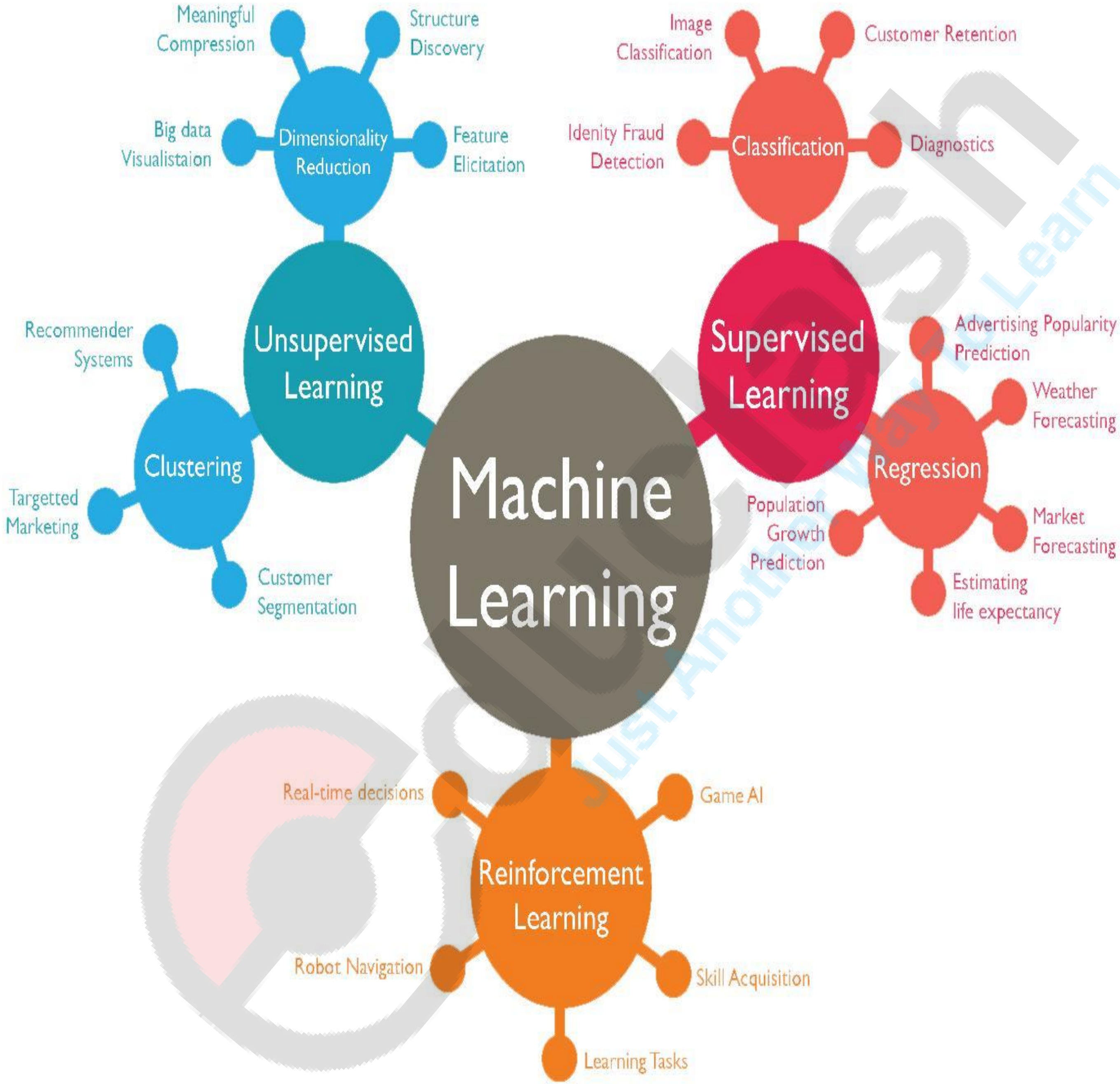


Weight	Wingspan	Webbed feet?	Back color	Species
1000.1	125.0	No	Brown	Buteo jamaicensis
3000.7	200.0	No	Gray	Sagittarius serpentarius

Figure 1.2 Features and target variable identified

Key tasks of machine learning

- Association
- Supervised Learning
  - Classification
  - Regression
- Unsupervised Learning
- Reinforcement Learning


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Key tasks of machine learning

Supervised learning tasks	
k-Nearest Neighbors	Linear
Naive Bayes	Locally weighted linear
Support vector machines	Ridge
Decision trees	Lasso
Unsupervised learning tasks	
k-Means	Expectation maximization
DBSCAN	Parzen window

Table 1.2 Common algorithms used to perform classification, regression, clustering, and density estimation tasks

How to choose the right algorithm

- If you're trying to predict or forecast a target value, then you need to look into supervised learning.
- discrete value like Yes/No, 1/2/3, A/B/C, or Red/Yellow/Black?
- If so, then you want to look into classification.
- If the target value can take on a number of values, say any value from 0.00 to 100.00, or -999 to 999, or + to -, then you need to look into regression.
- Else then unsupervised learning

How to choose the right algorithm

- trying to fit your data into some discrete groups?
- you should look into clustering.
- Do you need to have some numerical estimate of how strong the fit is into each group?
- If you answer yes, then you probably should look into a density estimation algorithm.

Steps in developing a machine learning application

- *Collect data.*
- *Prepare the input data.*
- *Analyze the input data.*
- *Train the algorithm.*
- *Test the algorithm*
- *Use it.*

Applications of ML - Learning Associations

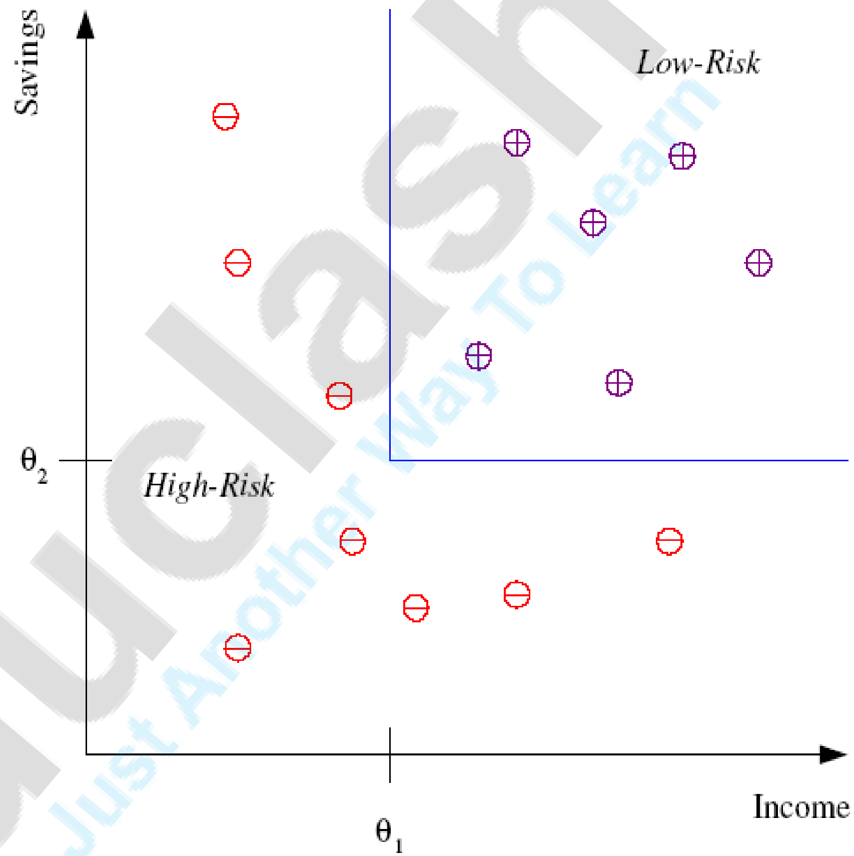
- Basket analysis:

  $P(Y \mid X)$ probability that somebody who buys X also buys Y where X and Y are products/services.

  Example: P (chips | beer) = 0.7

Applications of ML Classification

- Example: Credit scoring
- Differentiating between low-risk and high-risk customers from their *income* and *savings*



Discriminant: IF *income* > θ_1 AND *savings* > θ_2
THEN low-risk ELSE high-risk

Classification: Applications

- Aka Pattern recognition
- Face recognition: Pose, lighting, occlusion (glasses, beard), make-up, hair style
- Character recognition: Different handwriting styles.
- Speech recognition: Temporal dependency.
- Medical diagnosis: From symptoms to illnesses
- Biometrics: Recognition/authentication using physical and/or behavioral characteristics: Face, iris, signature, etc
- ...

Face Recognition

Training examples of a person

Test images

ORL dataset,
AT&T Laboratories, Cambridge UK

Regression

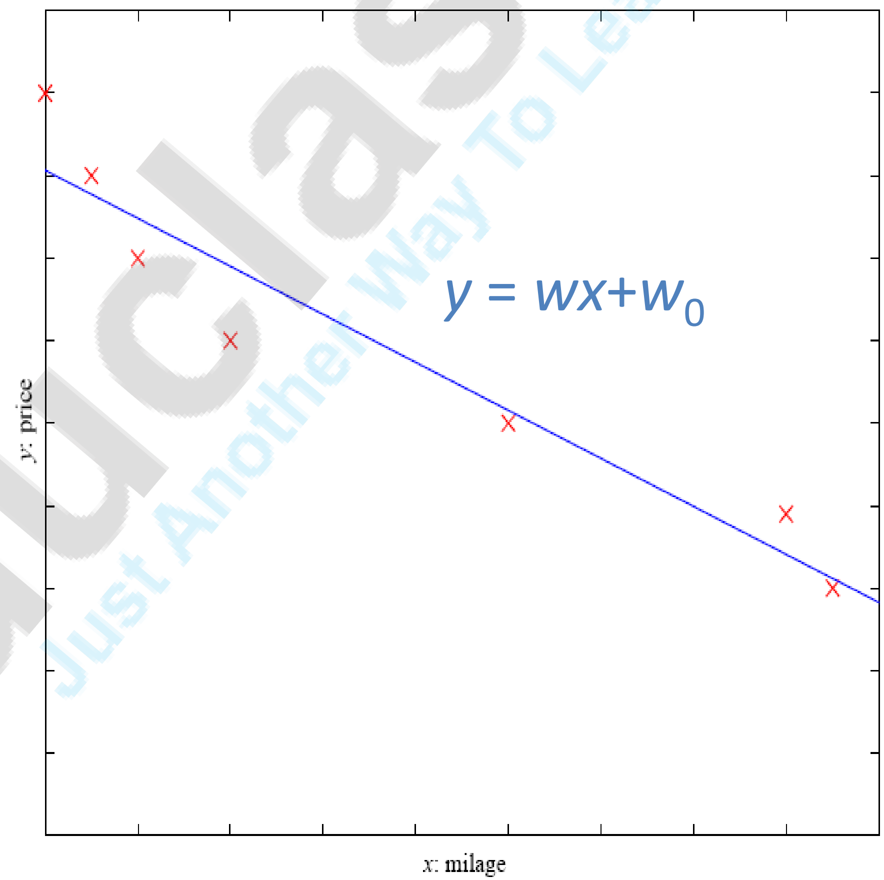
- Example: Price of a used car
- x : car attributes

  y : price

  $$y = g\,(x \mid \theta\,)$$

  g ( ) model,

  θ parameters



Supervised Learning: Uses

- Prediction of future cases: Use the rule to predict the output for future inputs
- Knowledge extraction: The rule is easy to understand
- Compression:
- Outlier detection: Exceptions that are not covered by the rule, e.g., fraud

Unsupervised Learning

- Learning “what normally happens”
- No output
- Clustering: Grouping similar instances
- Example applications
  - Customer segmentation in CRM
  - Image compression: Color quantization
  - Bioinformatics: Learning motifs

Reinforcement Learning

- Learning a policy: A sequence of outputs
- No supervised output but delayed reward
- Credit assignment problem
- Game playing
- Robot in a maze
- Multiple agents, partial observability, ...

Software Tools – Machine Learning

- Scikit-learn Developers. Scikit-learn.
- Super Data Science
- The Shogun Team. Shogun.
- Accord.Net Framework. Project: Accord Framework/AForge.net
- **Apache Singa :** *http://singa.apache.org/en/index.html*
- Apache **Software** Foundation. Apache Mahout
- Apache **Software** Foundation. Spark Mllib
  - practical machine learning scalable and easy. (all algms) *http://spark.apache.org/mllib/*
- **TensorFlow by Google**
- Cloudera Oryx2 built on Apache Spark and Apache Kafka for real-time large scale machine learning.
- **Amazon Machine Learning (AML)** *https://aws.amazon.com/machine-learning/*

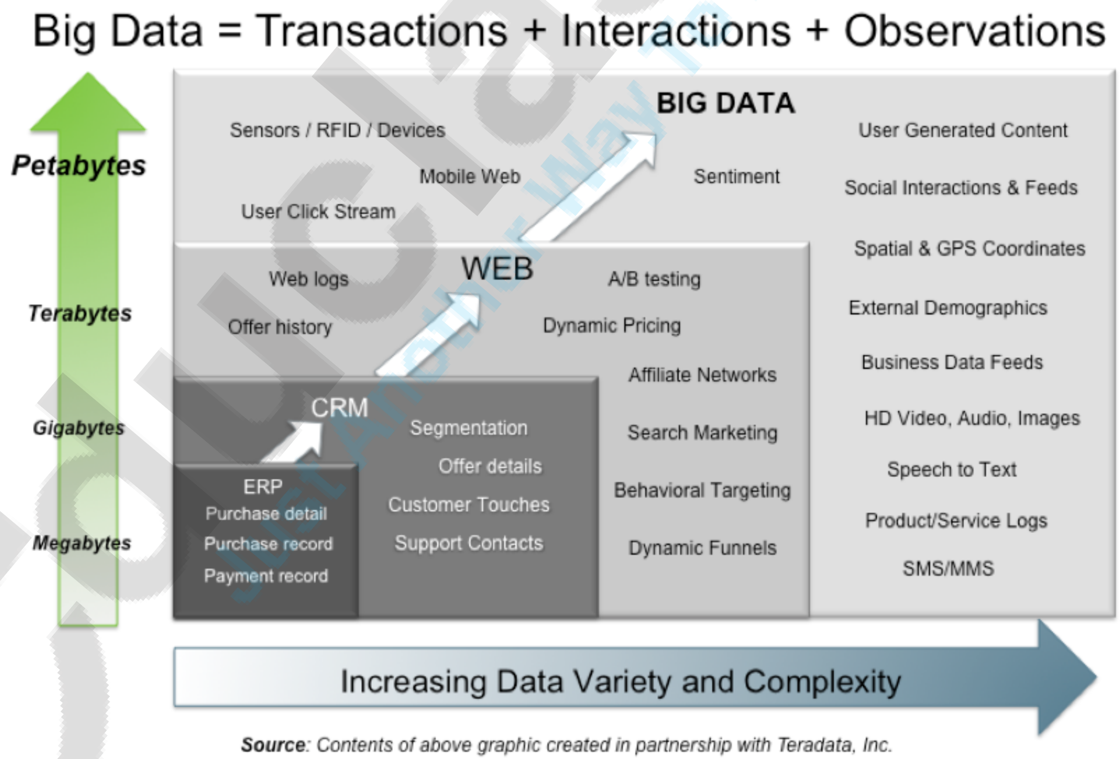
What’s Big Data?

No single definition; here is from Wikipedia:

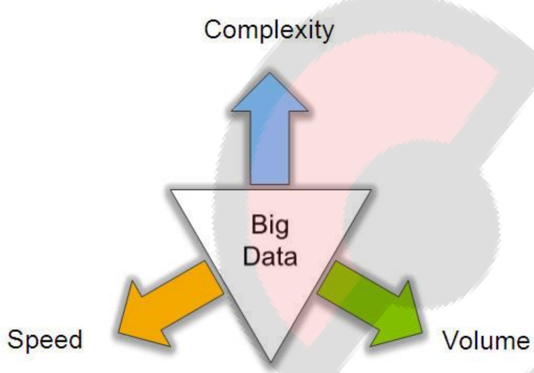
- **Big data** is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications.
- The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.
- The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, link legal citations, combat crime, and determine real-time roadway traffic conditions.”

Big Data: 3V's

BIG
DATA?
VOLUME
Large amounts of data.
VELOCITY
Needs to be analyzed quickly.
VARIETY
Different types of structured and unstructured data.



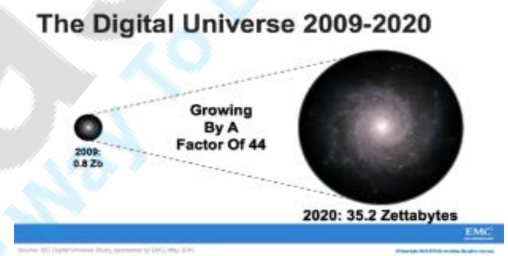
Big Data = Transactions + Interactions + Observations
Petabytes
Terabytes
Gigabytes
Megabytes
BIG DATA
Sensors / RFID / Devices
Mobile Web
Sentiment
User Click Stream
User Generated Content
Social Interactions & Feeds
Spatial & GPS Coordinates
External Demographics
Business Data Feeds
HD Video, Audio, Images
Speech to Text
Product/Service Logs
SMS/MMS
WEB
Web logs
Offer history
A/B testing
Dynamic Pricing
Affiliate Networks
Search Marketing
Behavioral Targeting
Dynamic Funnels
CRM
Segmentation
Offer details
Customer Touches
Support Contacts
ERP
Purchase detail
Purchase record
Payment record
Increasing Data Variety and Complexity
Source: Contents of above graphic created in partnership with Teradata, Inc.

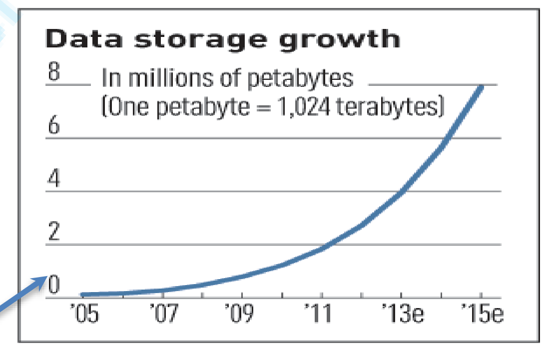


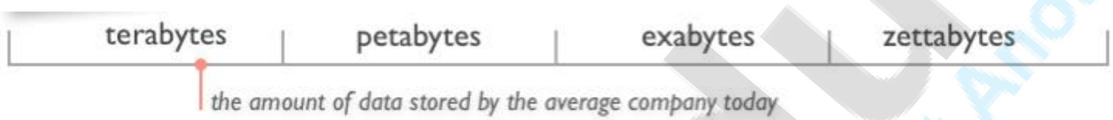
Complexity
Big Data
Speed
Volume

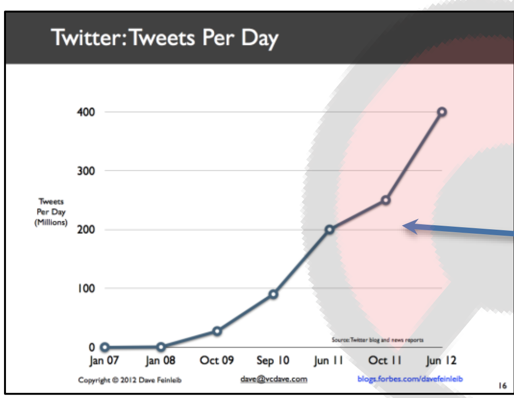
Volume (Scale)

- **Data Volume**
  - 44x increase from 2009 2020
  - From 0.8 zettabytes to 35zb
- Data volume is increasing exponentially





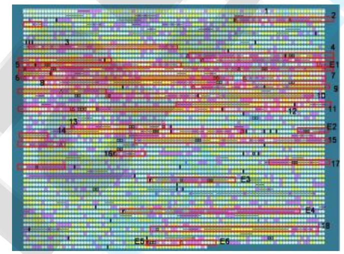


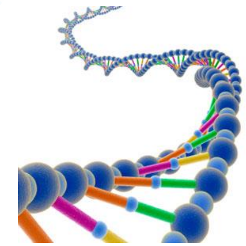


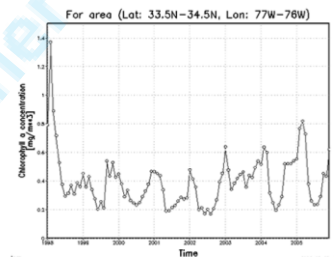
Exponential increase in collected/generated data

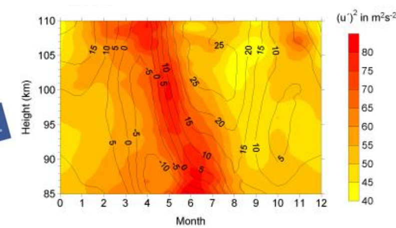
Variety (Complexity)

- Relational Data (Tables/Transaction/Legacy Data)
- Text Data (Web)
- Semi-structured Data (XML)
- Graph Data
  - Social Network, Semantic Web (RDF), …
- Streaming Data
  - You can only scan the data once
- A single application can be generating/collecting many types of data
- Big Public Data (online, weather, finance, etc)









To extract knowledge➔ all these types of data need to linked together

Velocity (Speed)

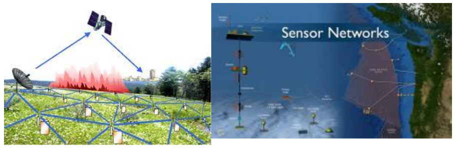
- Data is begin generated fast and need to be processed fast
- Online Data Analytics
- Late decisions ➔ missing opportunities
- **Examples**
  - **E-Promotions:** Based on your current location, your purchase history, what you like ➔ send promotions right now for store next to you
  - **Healthcare monitoring:** sensors monitoring your activities and body ➔ any abnormal measurements require immediate reaction

Real-time/Fast Data

Social media and networks
(all of us are generating data)

Scientific instruments
(collecting all sorts of data)

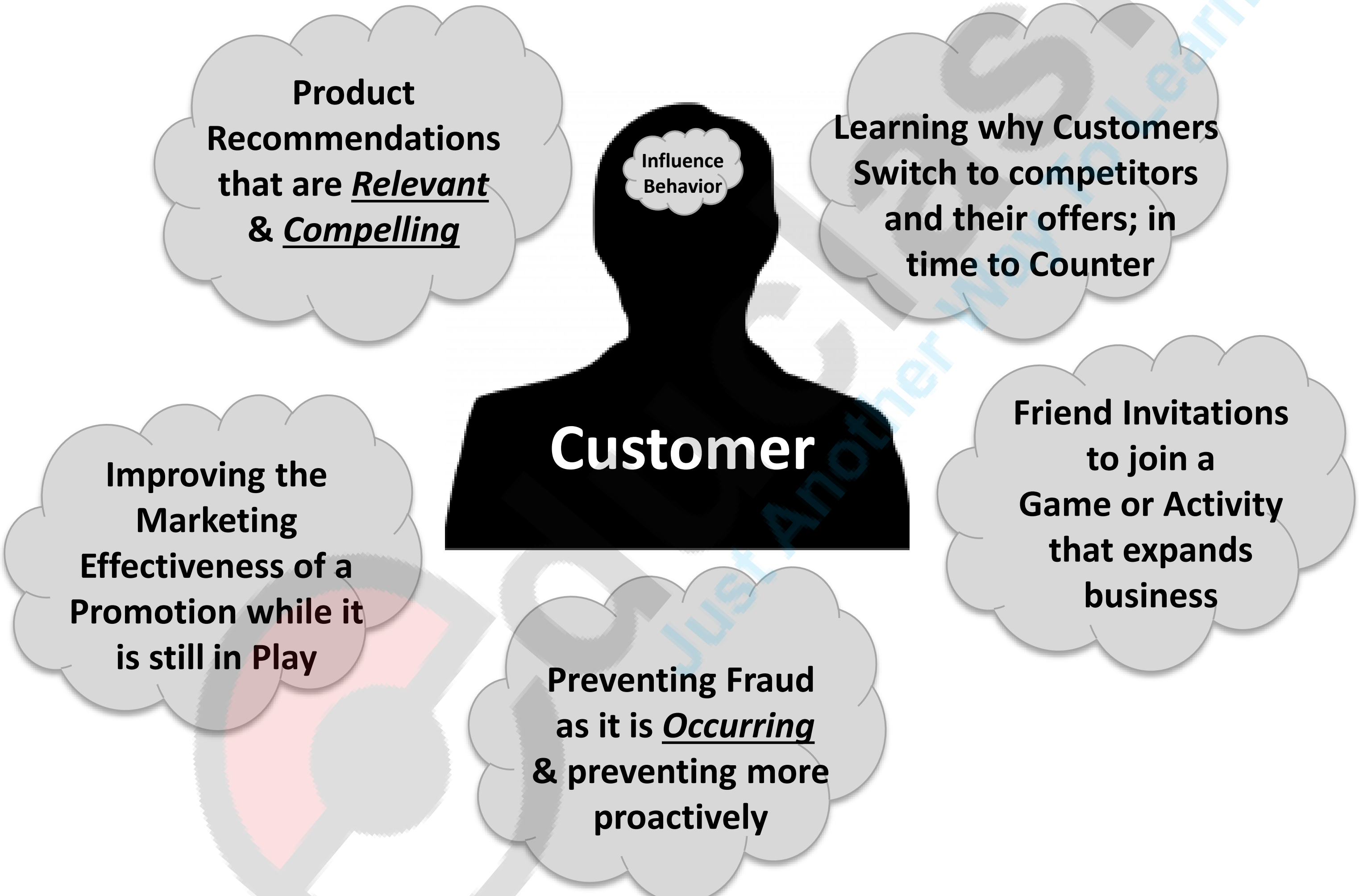
Mobile devices
(tracking all objects all the time)



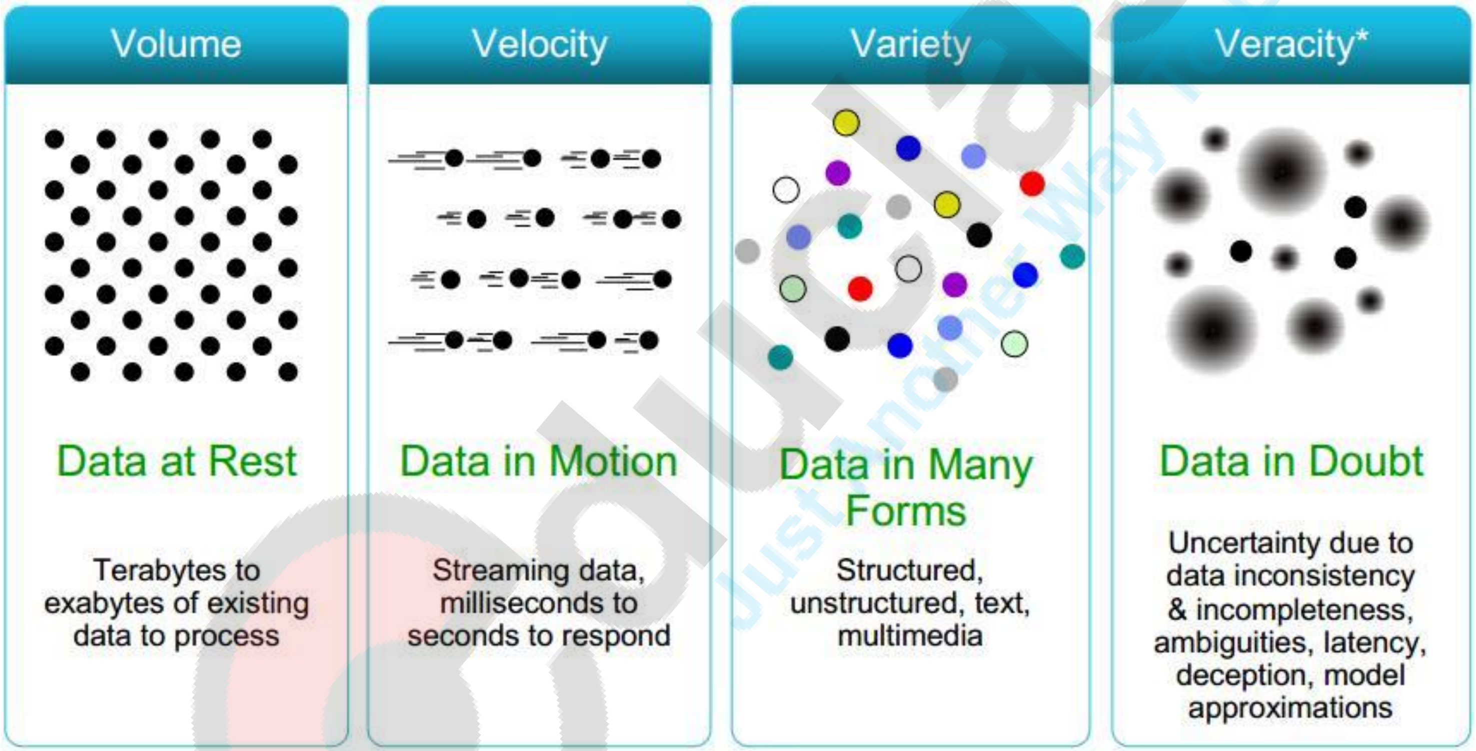
Sensor technology and networks
(measuring all kinds of data)

- The progress and innovation is no longer hindered by the ability to collect data
- But, by the ability to manage, analyze, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion

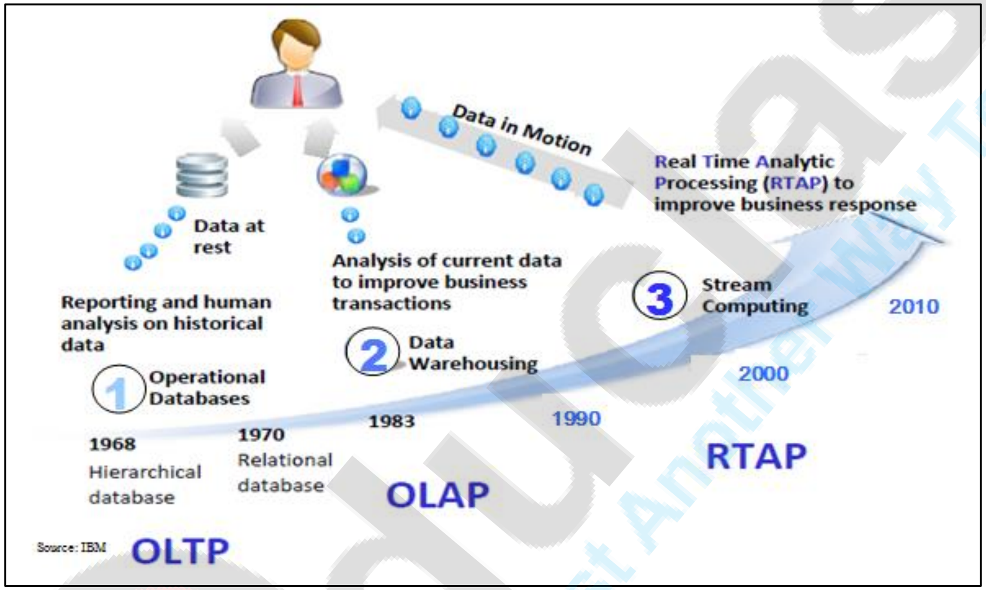
Real-Time Analytics/Decision Requirement



Some Make it 4V's



Harnessing Big Data



- **OLTP:** Online Transaction Processing (DBMSs)
- **OLAP:** Online Analytical Processing (Data Warehousing)
- **RTAP:** Real-Time Analytics Processing (Big Data Architecture & technology)

The Model Has Changed…

- **The Model of Generating/Consuming Data has Changed**

Old Model: Few companies are generating data, all others are consuming data

New Model: all of us are generating data, and all of us are consuming data

Big Data and Machine Learning

- Big data analytics is the process of collecting and analyzing the large volume of data sets (called Big Data) to discover useful hidden patterns and other information like customer choices, market trends that can help organizations make more informed and customer oriented business decisions.
- Big data is a term that describes the data characterized by 3Vs: the extreme volume of data, the wide variety of data types and the velocity at which the data must be processed.
- Big data can be analyzed for insights that lead to better decisions and [strategic] business moves.

Machine Learning

- Machine learning is a field of AI (Artificial Intelligence)
- to increase their accuracy for the expecting outcomes.
  - You know those movie/show recommendations you get on Netflix or Amazon? Machine learning does this for you.
  - How does Uber/Ola determine the price of your cab ride? How do they minimize the wait time once you hail a car? How do these services optimally match you with other passengers to minimize detours? The answer to all these questions is Machine Learning.
  - How can a financial institution determine if a transaction is fraudulent or not? In most cases, it is difficult for humans to manually review each transaction because of its very high daily transaction volume. Instead, AI is used to create systems that learn from the available data to check what types of transactions are fraudulent.
  - Ever wondered what's the technology behind the self-driving Google car? Again the answer is machine learning.
- Now we know What Big Data vs Machine Learning are, but to decide which one to use at which place we need to see the difference between both.

#1. Data Use

Big Data

Big data can be used for a variety of purposes, including financial research, collecting sales data etc.

Machine Learning

Machine learning is the technology behind self-driving cars and advance recommendation engines.

#2. Foundations for Learning

Big Data

Big data analytics pulls from existing information to look for emerging patterns that can help shape our decision-making processes.

Machine Learning

On the other hand, Machine learning can learn from the existing data and provide the foundation required for a machine to teach itself.

#3. Pattern Recognition

Big Data	Machine Learning
Big data analytics can reveal some patterns through classifications and sequence analysis.	However, machine learning takes this concept a one step ahead by using the same algorithms that big data analytics uses to automatically learn from the collected data.

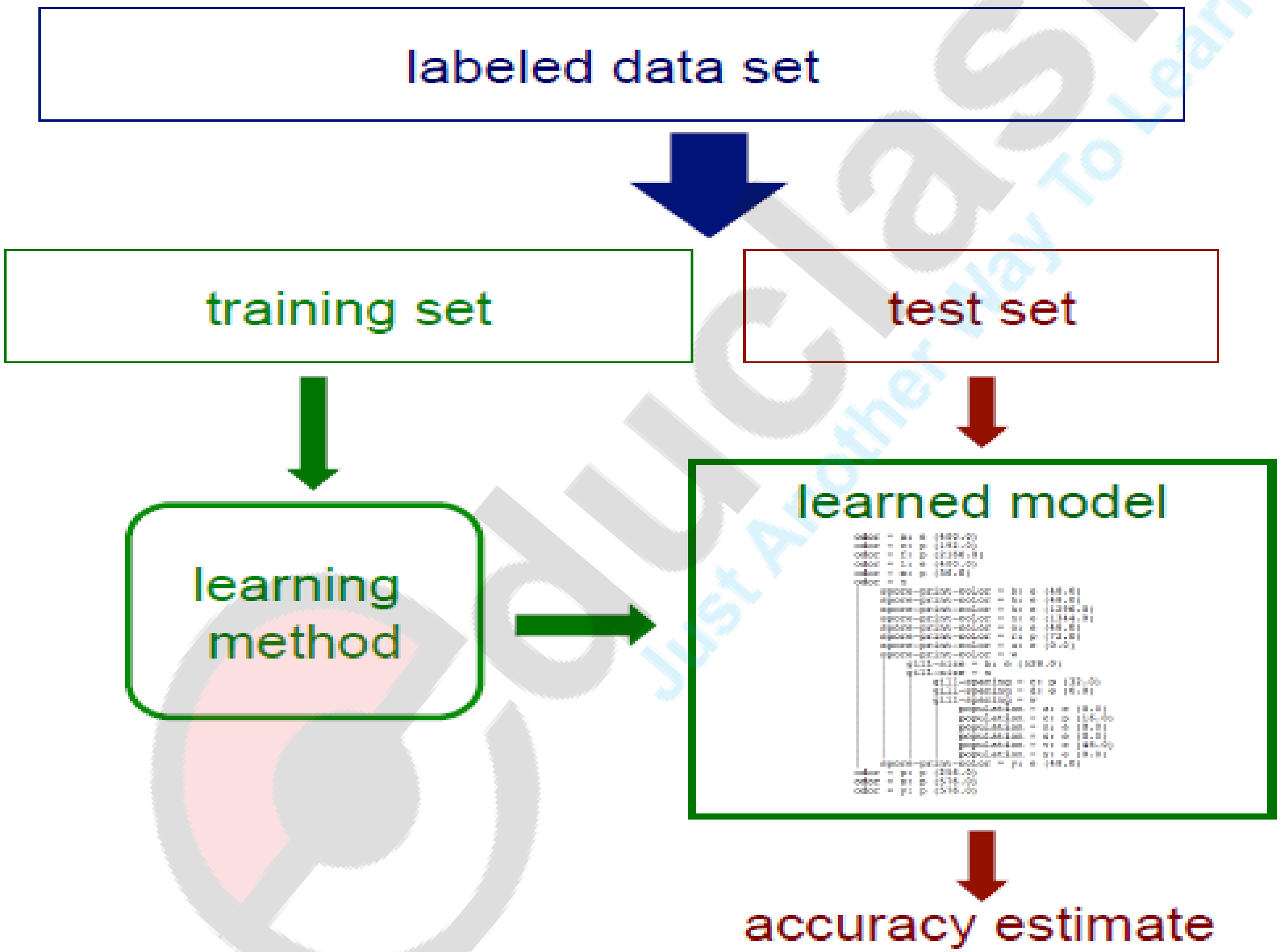
#4. Data Volume

Big Data	Machine Learning
Big data as the name suggest tends to be interested in large-scale datasets where the problem is dealing with the large volume of data.	ML tends to be more interested in small datasets where over-fitting is the problem.

#5. Purpose

Big Data	Machine Learning
Purpose of big data is to store large volume of data and find out pattern in data .	Purpose of machine learning is to learn from trained data and predicts or estimates future results.

Hypothesis space



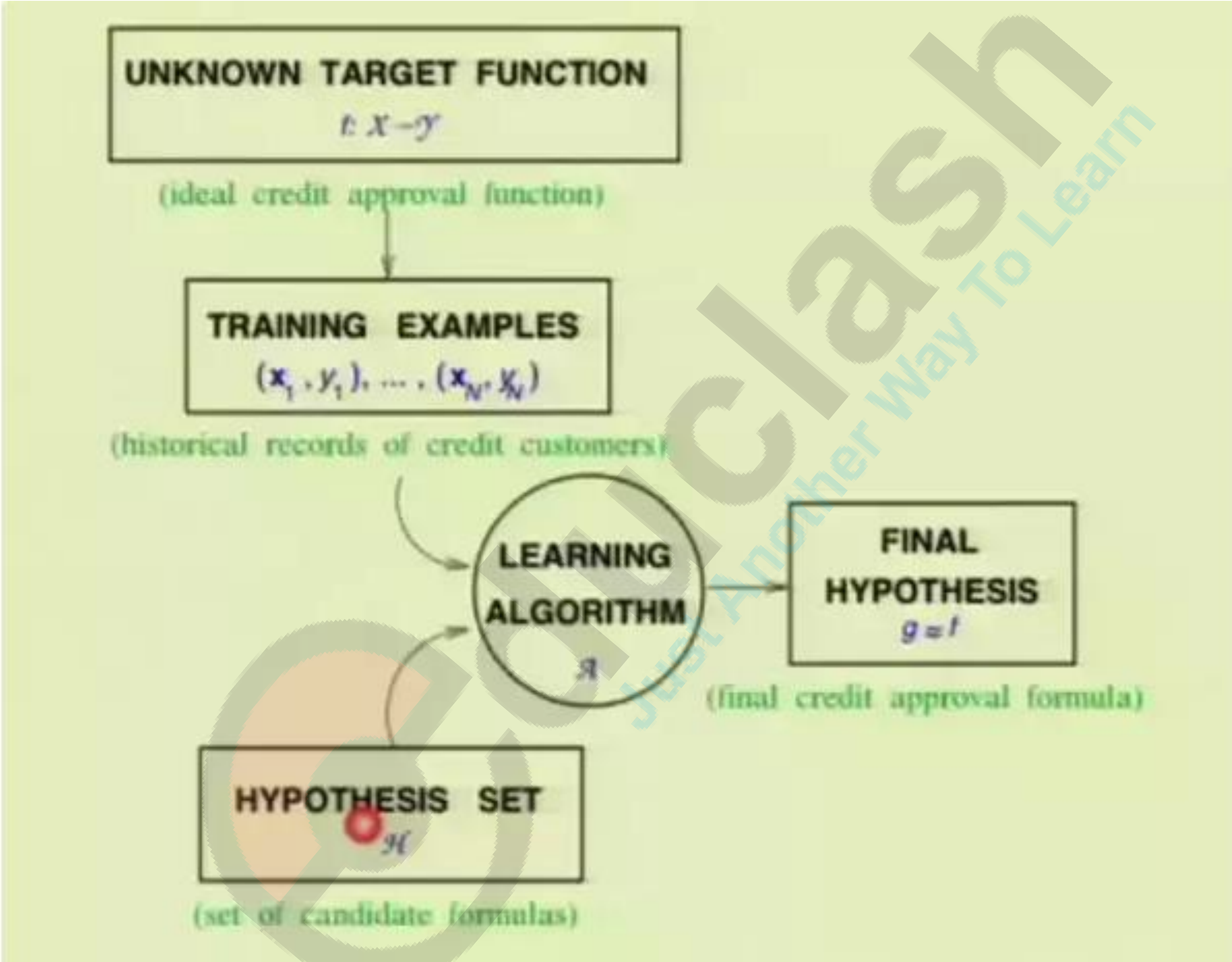
Hypothesis Space Example

- Example
- Suppose an example with four binary features and one binary output variable. Below is a set of observations:
- x1 x2 x3 x4 | y ---------------
- 0 0 0 1 | 0
- 0 1 0 1 | 0
- 1 1 0 0 | 1
- 0 0 1 0 | 1

- This set of observations can be used by a **machine learning (ML) algorithm** to learn a function **f** that is able to predict a value y for any input from the **input space**.
- We are searching for the ground truth f(x) = y that explains the relation between x and y for all possible inputs in the correct way.
- The function f has to be chosen from the **hypothesis space**.
- To get a better idea:
- The input space is in the above given example , its the number of possible inputs.
- The hypothesis space is because for each set of features of the input space two outcomes (0 and 1) are possible.
- The ML algorithm helps us to find **one function**, sometimes also referred as hypothesis, from the relatively large hypothesis space.

- In a machine learning problem where the input is denoted by x and the output is y
- In order to do machine learning, there should exist a relationship (pattern) between the input and output values. Lets say that this the function y=f(x), this known as the **target function.**
-

  However, f(.) is unknown function to us. so machine learning algorithms try to guess a `hypothesis' function that approximates the unknown f(.)
- the set of all possible hypotheses is known as the Hypothesis set H(.)
- the goal is the learning process is to find the final hypothesis that best approximates the unknown target function.
- Hypothesis set may include linear formula, neural net function, support vector machine. And the learning algorithm include backprogation, gradient descent.


UNKNOWN TARGET FUNCTION
$f: \mathcal{X} \to \mathcal{Y}$
(ideal credit approval function)
TRAINING EXAMPLES
$(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$
(historical records of credit customers)
LEARNING ALGORITHM
$\mathcal{A}$
FINAL HYPOTHESIS
$g \approx f$
(final credit approval formula)
HYPOTHESIS SET
$\mathcal{H}$
(set of candidate formulas)

Estimate hypothesis accuracy

- Evaluating the performance of learning systems is important because:
  - Learning systems are usually designed to predict the class of "future" unlabeled data points.
  - In some cases, evaluating hypotheses is an integral part of the learning process (example, when pruning a decision tree)

Difficulties in Evaluating Hypotheses when only limited data are available

- ***Bias in the estimate:*** The observed accuracy of the learned hypothesis over the training examples is a poor estimator of its accuracy over future examples ==> we test the hypothesis on a test set chosen independently of the training set and the hypothesis.
- ***Variance in the estimate:*** Even with a separate test set, the measured accuracy can vary from the true accuracy, depending on the makeup of the particular set of test examples. The smaller the test set, the greater the expected variance.

Questions Considered

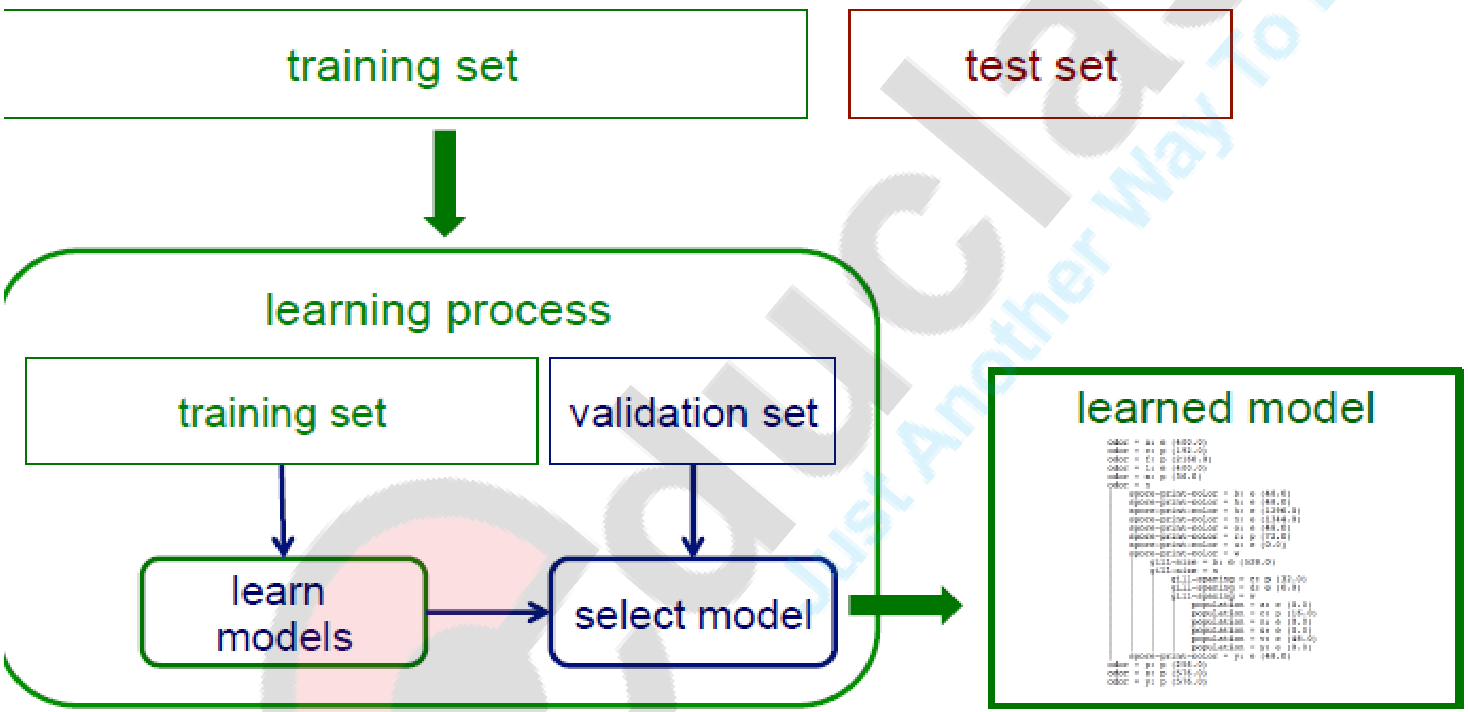
- Given the observed accuracy of a hypothesis over a limited sample of data, how well does this estimate its accuracy over additional examples?
- Given that one hypothesis outperforms another over some sample data, how probable is it that this hypothesis is more accurate, in general?
- When data is limited what is the best way to use this data to both learn a hypothesis and estimate its accuracy?

Is accuracy an adequate measure of predictive performance?

- accuracy may not be useful measure in cases where
  - there is a large class skew
    - Is 98% accuracy good if 97% of the instances are negative?
  - there are differential misclassification costs – say, getting a positive wrong costs more than getting a negative wrong
    - Consider a medical domain in which a false positive results in an extraneous test but a false negative results in a failure to treat a disease
  - we are most interested in a subset of high-confidence predictions

Validation (tuning) sets revisited

Suppose we want unbiased estimates of accuracy during the learning process (e.g. to choose the best level of decision-tree pruning)?



Partition training data into separate training/validation sets

Limitations of using a single training/test partition

- we may not have enough data to make sufficiently large training and test sets
  - a larger test set gives us more reliable estimate of accuracy (i.e. a lower variance estimate)
  - but… a larger training set will be more representative of how much data we actually have for learning process

- a single training set doesn't tell us how sensitive accuracy is to a particular training sample

Is accuracy an adequate measure of predictive performance?

- accuracy may not be useful measure in cases where
  - there is a large class skew
    - Is 98% accuracy good if 97% of the instances are negative?
  - there are differential misclassification costs – say, getting a positive wrong costs more than getting a negative wrong
    - Consider a medical domain in which a false positive results in an extraneous test but a false negative results in a failure to treat a disease
  - we are most interested in a subset of high-confidence predictions

Confusion matrix for 2-class problems

		actual class	
		positive	negative
predicted class	positive	true positives (TP)	false positives (FP)
	negative	false negatives (FN)	true negatives (TN)

$$\text{accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

CONFUSION MATRIX

True Positive (TP): • Reality: A wolf threatened. • Shepherd said: "Wolf." • Outcome: Shepherd is a hero.	**False Positive (FP):** • Reality: No wolf threatened. • Shepherd said: "Wolf." • Outcome: Villagers are angry at shepherd for waking them up.
False Negative (FN): • Reality: A wolf threatened. • Shepherd said: "No wolf." • Outcome: The wolf ate all the sheep.	**True Negative (TN):** • Reality: No wolf threatened. • Shepherd said: "No wolf." • Outcome: Everyone is fine.

- A true positive is an outcome where the model correctly predicts the positive class.
- A true negative is an outcome where the model correctly predicts the negative class.
- A false positive is an outcome where the model incorrectly predicts the positive class.
- A false negative is an outcome where the model incorrectly predicts the negative class.

ACCURACY

True Positive (TP): • Reality: Malignant • ML model predicted: Malignant • Number of TP results: 1	**False Positive (FP):** • Reality: Benign • ML model predicted: Malignant • Number of FP results: 1
False Negative (FN): • Reality: Malignant • ML model predicted: Benign • Number of FN results: 8	**True Negative (TN):** • Reality: Benign • ML model predicted: Benign • Number of TN results: 90

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{1+90}{1+90+1+8} = 0.91$$

Accuracy comes out to 0.91, or 91% (91 correct predictions out of 100 total examples). Of the 100 tumor examples, 91 are benign (90 TNs and 1 FP) and 9 are malignant (1 TP and 8 FNs).

ACCURACY

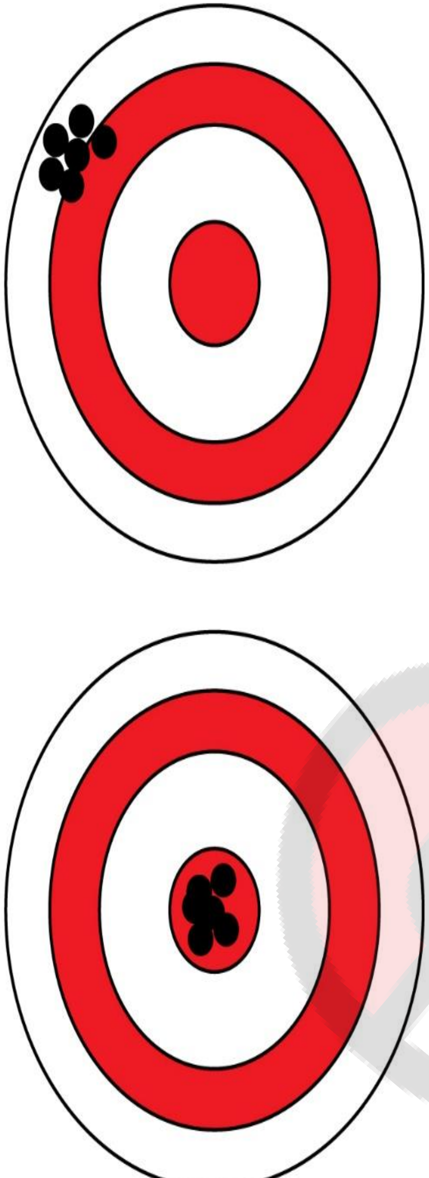
True Positive (TP): • Reality: Malignant • ML model predicted: Malignant • Number of TP results: 1	False Positive (FP): • Reality: Benign • ML model predicted: Malignant • Number of FP results: 1
False Negative (FN): • Reality: Malignant • ML model predicted: Benign • Number of FN results: 8	True Negative (TN): • Reality: Benign • ML model predicted: Benign • Number of TN results: 90

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{1 + 90}{1 + 90 + 1 + 8} = 0.91$$

Of the 91 benign tumors, the model correctly identifies 90 as benign. That's good. However, of the 9 malignant tumors, the model only correctly identifies 1 as malignant—a terrible outcome, as 8 out of 9 malignancies go undiagnosed!

While 91% accuracy may seem good at first glance, another tumor-classifier model that always predicts benign would achieve the exact same accuracy (91/100 correct predictions) on our examples. In other words, our model is no better than one that has zero predictive ability to distinguish malignant tumors from benign tumors.

ACCURACY VS. PRECISION



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

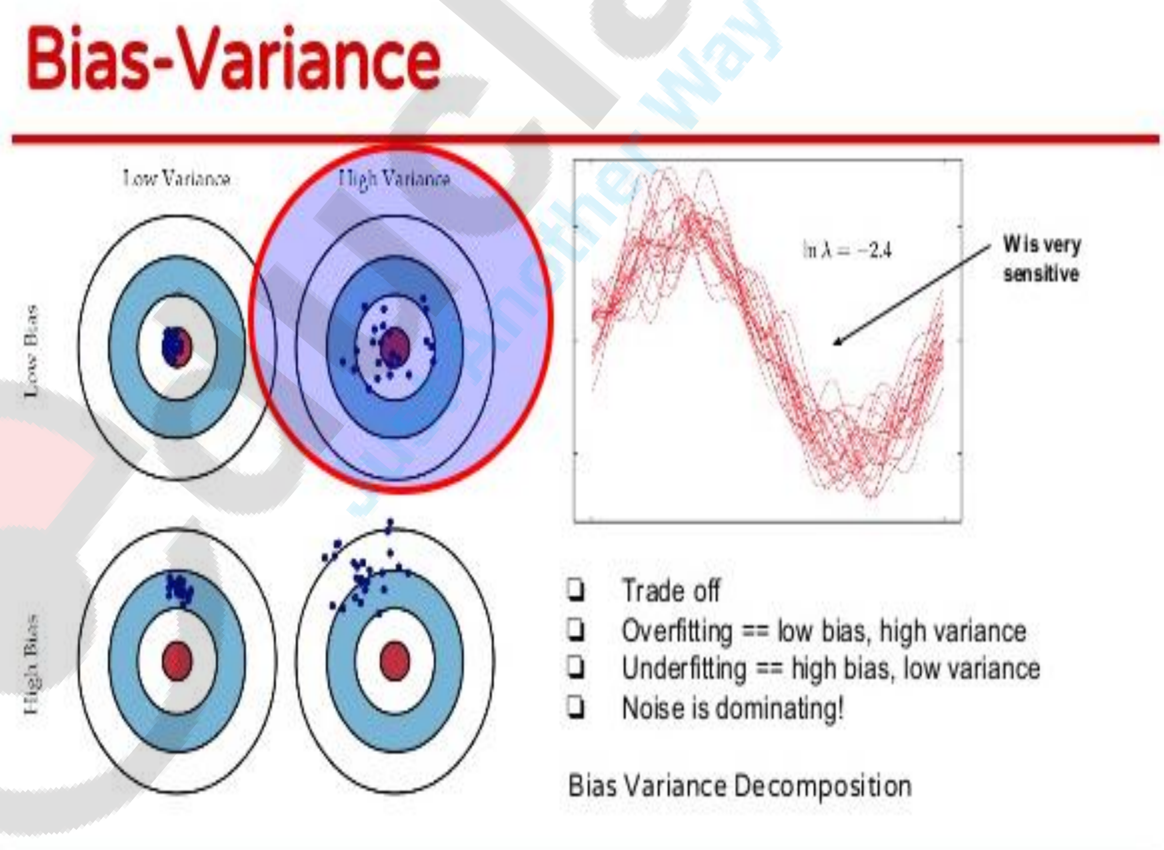
$$\text{Precision} = \frac{TP}{TP + FP}$$

ERROR

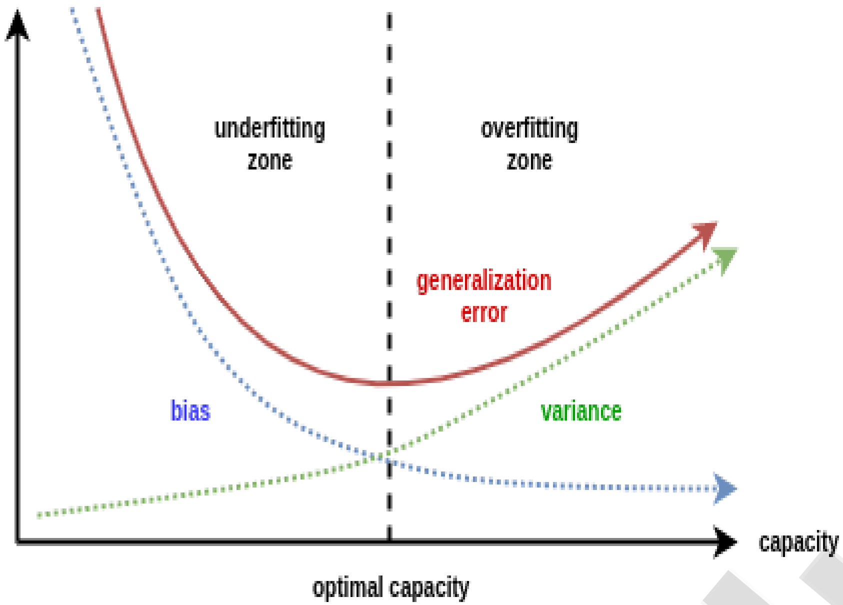
Error is inherent is any modeling. The error emerging from any model can be broken down into three components:

$$Err(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\hat{f}(x) - E[\hat{f}(x)]\right]^2 + \sigma_e^2$$

$$Err(x) = \mathrm{Bias}^2 + \mathrm{Variance} + \mathrm{Irreducible\ Error}$$



BIAS VS VARIANCE

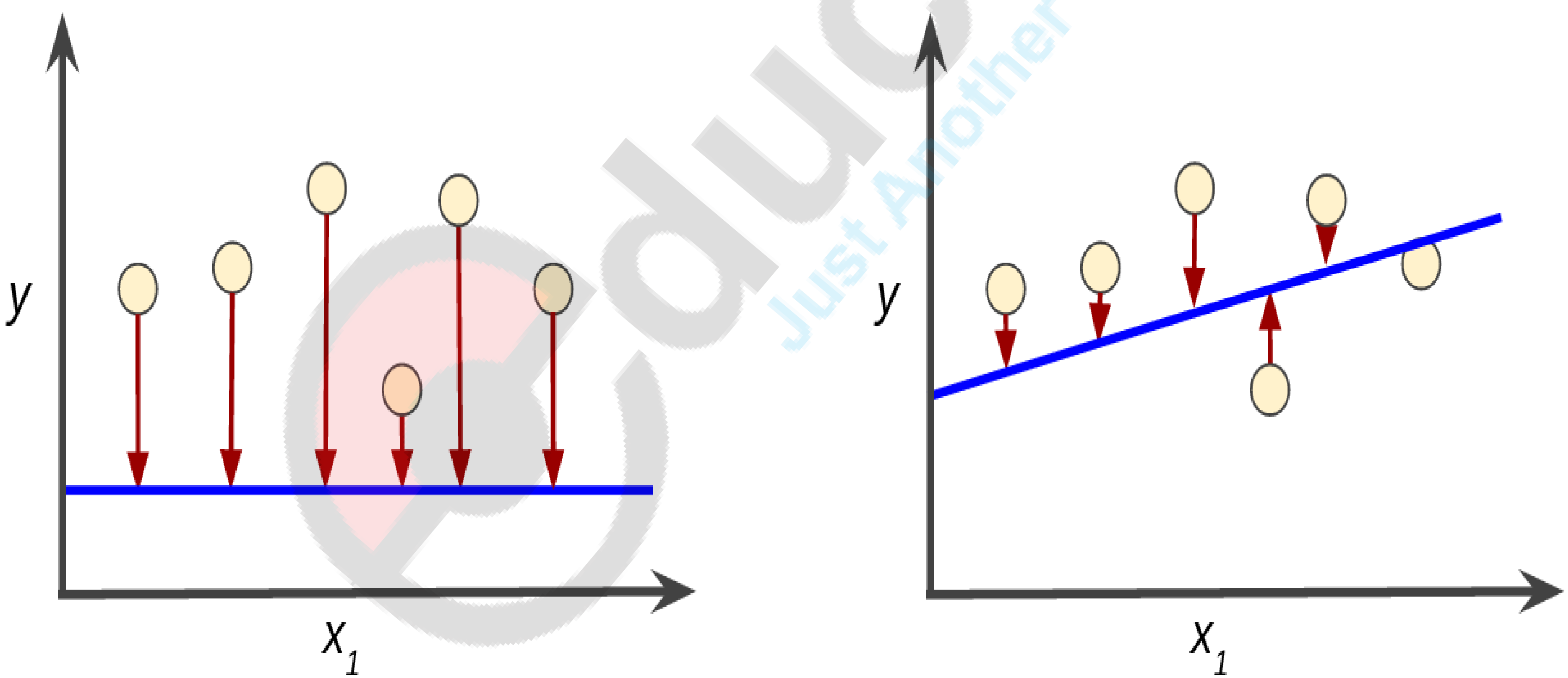


Bias and variance measure two different sources of an estimator's error: bias measures the expected deviation from the true value of the function or parameter, and variance measures the deviation from the expected estimator value that any given sampling of from the data generating distribution is likely to cause.

One way to address this trade-off is by using cross-validation, in which the training data is partitioned into k equally subsets, each of which is "held out" to use as validation data in a series of training, validation "rounds". The results from evaluating the estimator on the validation data in each round is averaged to produce a estimated generalization error. Alternatively, we may use the mean squared error (MSE) to compare estimators:

LOSS

Loss is the penalty for a bad prediction. That is, loss is a number indicating how bad the model's prediction was on a single example. The goal of training a model is to find a set of weights and biases to minimize, if not eliminate, the loss.



LOSS

Mean squared error	$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^2$
Root mean squared error	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$
Mean absolute error	$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n} \lvert e_t \rvert$
Mean absolute percentage error	$\text{MAPE} = \frac{100\%}{n}\sum_{t=1}^{n} \left\lvert \frac{e_t}{y_t} \right\rvert$

The linear regression models typically use a loss function called squared loss (also known as L2 loss). The squared loss is $(y - y')^2$

Mean square error (MSE) is the average squared loss per example over the whole dataset. To calculate MSE, sum up all the squared losses for individual examples and then divide by the number of examples:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2$$

where N is the number of data points, f_i the value returned by the model and y_i the actual value for data point i.

Hypothesis testing

- Make Assumptions.
- Take an initial position. :**NULL Hypothesis (Ho)**
- Determine the alternate position. : **The Alternate Hypothesis (Ha)**
- Set acceptance criteria : a threshold needs to be set
- Conduct fact based tests.
- Evaluate results. Does the evaluation support the initial position? Are we confident that the result is not due to chance?
- Reach one of the following conclusion: Reject the original position in favor of alternate position or fail to reject the initial position.

educlash | Just Another Way To Learn

- **Join our Facebook group:** https://www.facebook.com/groups/educlashco
- **Like our Facebook Page:** https://www.facebook.com/educlashco
- **Join WhatsApp group:** https://educlash.com/join-educlash-com-whatsapp-group/
- **Follow Instagram:** https://www.Instagram.com/educlashco
- **Follow Twitter:** https://www.twitter.com/educlashco
- **Join Telegram:** https://t.me/educlashco